

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 22:10:41 ; Search time 41 Seconds
(without alignments)
502.921 Million cell updates/sec

Title: US-09-869-334b-1
Perfect score: 2052
Sequence: 1 MNVLNRQALQRLNGKNK.....SITPIENSVIYGLXSPRYVM 396

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt.42.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2052	100.0	396	1 YJIB_BACSU	O34374 bacillus su
2	839	40.9	405	1 CPXM_BACSU	P27632 bacillus su
3	738	36.0	410	1 CPXM_BACME	O06069 bacillus me
4	711.5	34.7	410	1 CPXI_BACME	P14762 bacillus me
5	658	32.1	410	1 CPXV_BACSU	O08469 bacillus su
6	646.5	31.5	405	1 CPXK_SACER	P33271 saccharopol
7	600	29.2	395	1 BHOI_BACSU	P53554 bacillus su
8	571	27.8	397	1 CPXQ_SACER	P48635 saccharopol
9	557	27.1	404	1 CPXJ_SACER	O00441 saccharopol
10	548.5	26.7	368	1 CPXW_SULSO	O55080 sulfolobus
11	520	25.3	428	1 C124_MYCTU	O50696 mycobacteri
12	516	25.1	402	1 C123_MYCTU	P77902 mycobacteri
13	514.5	25.1	398	1 C142_MYCTU	O53563 mycobacteri
14	499	24.3	433	1 C125_MYCTU	P71856 mycobacteri
15	489.5	23.9	402	1 C13B_XYLFA	O9965 xylella fas
16	488	23.8	414	1 C126_MYCTU	P77903 mycobacteri
17	486.5	23.7	402	1 C13E_XYLFT	O87ax5 xylella fas
18	482	23.5	406	1 CPXO_PSRPU	O59723 pseudomonas
19	475	23.1	405	1 CPXV_STRGO	P18326 streptomyc
20	474	23.1	400	1 C13C_XYLFA	O9966 xylella fas
21	471.5	23.0	438	1 C140_MYCTU	O08464 mycobacteri
22	465	22.7	399	1 C13C_XYLFT	O87av9 xylella fas
23	462	22.5	402	1 CPXF_STRGO	P18327 streptomyc
24	451	22.0	405	1 C130_MYCTU	Q11062 mycobacteri
25	446	21.7	381	1 CPXG_STRSQ	P23296 streptomyc
26	439	21.4	409	1 CP52_STRCC	O59831 streptomyc
27	437.5	21.3	402	1 NOR_FUSOX	P23295 fusarium ox
28	435	20.7	399	1 PASI_RHOFA	P46373 rhodococcus
29	424	20.7	412	1 CPXH_STRGR	P26311 streptomyc
30	419.5	20.4	428	1 CPXJ_PSSSP	P33006 pseudomonas
31	418.5	20.4	405	1 CPXK_BACSU	O34926 bacillus su
32	418.5	20.4	406	1 CSC4_AMTOR	O8rn03 amycolatops
33	415	20.2	408	1 NOR2_CYLTO	Q12599 cylindrocac

34	403	19.6	403	1	NOR1_CYLTO	Q00616 cylindrocac
35	401.5	19.6	398	1	C5B3_AMTOR	O8rn04 amycolatops
36	397	19.3	401	1	CPXP_RHISN	P55544 rhizobium s
37	391	19.1	401	1	CPXP_BRAJA	O59203 bradyrhizob
38	390	19.0	436	1	THCB_RHOBR	P43492 rhodococcus
39	385	18.8	489	1	C128_MYCTU	O59572 mycobacteri
40	376	18.3	400	1	C141_MYCTU	O08362 mycobacteri
41	375.5	18.3	429	1	CPXR_BRAJA	O59204 bradyrhizob
42	366.5	17.9	414	1	YAVG_RHISN	Q53215 rhizobium s
43	362	17.6	414	1	CPXA_PSRPU	P00183 pseudomonas
44	355	17.3	391	1	C5A3_AMTOR	O8rn05 amycolatops
45	351	17.1	422	1	CPXC_AGRTU	P24466 agrobacteri

ALIGNMENTS

RESULT 1
YJIB_BACSU
ID YJIB_BACSU STANDARD; PRT; 396 AA.
AC O34374;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative cytochrome P450 YJIB (EC 1.14.-.-).
GN YJIB OR BSU12210.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID:1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Rivoita C., Soldo B., Lazarevic V., Joris B., Maue C., Karamata D.;
RT "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a
RT putative 12.3 kb operon involved in hexuronate catabolism and a
RT perfect catabolite-responsible element."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot P., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guilseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpe P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassartotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

CC	-----	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	-----	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	-----	the European Bioinformatics Institute. There are no restrictions on its
CC	-----	use by non-profit institutions as long as its content is in no way
CC	-----	modified and this statement is not removed. Usage by and for commercial
CC	-----	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	-----	or send an email to license@isb-sib.ch).
CC	-----	-----
CC	-----	EMBL; AF015825; AAC46317.1; --
DR	EMBL; Z99110; CAB13078.1; --	
DR	PIR; B69851; B69851.	
DR	HSSP; Q55080; 1109.	
DR	Subtilist; BG13195; vJiB.	
DR	InterPro; IPR001128; Cytochrome_P450.	
DR	Pfam; PF00067; P450; 2.	
DR	PRINTS; PR00385; P450.	
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.	
DR	Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;	
KW	Complete proteome.	
KW	METAL 349 349	
FT	SEQUENCE 396 AA; 449990 MW; 9A89CF12613DBCFF CRC64;	
ST	IRON (HEME AXIAL LIGAND) (BY SIMILARITY).	
SO		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 18:12:54 ; Search time 3275 Seconds
(without alignments)
15762.306 Million cell updates/sec

Title: US-09-869-334B-2
Perfect score: 1191
Sequence: 1 atgaatgtgtaaacgcgcg.....gcttcgcgtgaaatgtaa 1191

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_bai.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_scs.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_scs.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1189.4	99.9	35739	1	AF015825	AF015825 Bacillus
2	1189.4	99.9	201241	1	BSUB00007	Z99110 Bacillus su
3	425	35.7	966	6	AX433775	AX433775 Sequence
4	242	20.3	1213	6	AX433767	AX433767 Sequence
5	234.4	19.7	290685	1	AE017032	AE017032 Bacillus
6	229.8	19.3	3187	1	BACRTP	M24523 B.subtilis
7	182.6	15.3	4317	1	BMCTP450A	Z11972 B.megateriu
8	169	14.2	1902	1	BMCP4501	X16610 Bacillus me
9	165.8	13.9	6629	1	AB072568	AB072568 Streptomy
10	162.4	13.6	5008	1	BSCYP4ZL	Y11043 Bacillus su
11	162.4	13.6	17828	1	BSUB93876	U93876 Bacillus su
12	162.4	13.6	19173	1	BSUB0014	Z99117 Bacillus su
13	159.6	13.4	300893	1	AE017006	AE017006 Bacillus
14	156.4	13.1	10167	1	SFU08223	U08223 Streptomyce
15	150	12.6	342300	1	MLEPRTN8	AL583924 Mycobacte
16	149	12.5	1233	6	AX697983	AX697983 Sequence
17	149	12.5	60196	6	AX697977	AX697977 Sequence
18	145.8	12.2	17512	1	AB071405	AB071405 Lechevali
19	145.8	12.2	25681	1	SAE414559	AJ414559 Saccharot
20	145.8	12.2	26144	1	AB090952	AB090952 Lechevali
21	145.8	12.2	28654	1	AF534707	AF534707 Lechevali
22	144.2	12.1	3078	1	SERC450A	M83110 Saccharopol
23	144	12.1	39228	1	MLCB1788	AL008609 Mycobacte
24	139.2	11.7	8478	6	AR173058	AR173058 Sequence
25	139.2	11.7	10153	1	BSU51868	U51868 Bacillus su
26	139.2	11.7	19922	1	BSUB0016	Z99119 Bacillus su
27	139.2	11.7	220060	1	AF008220	AF008220 Bacillus
28	138.2	11.6	207829	1	BSUB0010	Z99113 Bacillus su
29	136.4	11.5	6739	1	AB088066	AB088066 Bacillus
30	135.4	11.4	299175	1	AP005023	AP005023 Streptomy
31	125	10.5	300350	1	AP006574	AP006574 Gloebact
32	124.2	10.4	145911	1	AP003014	AP003014 Mesorhizo
33	122.4	10.3	309050	1	SCO939117	AL939117 Streptomy
34	120.6	10.1	10417	1	AE011961	AE011961 Xanthomon
35	120.6	10.1	300885	1	AE012559	AE012559 Xylella f
36	119.2	10.0	10029	1	AE009087	AE009087 Agrobacte
37	119.2	10.0	10184	1	AE008053	AE008053 Agrobacte
38	118.2	9.9	154620	2	EX323047	EX323047 Danio rer
39	118.2	9.9	16057	1	AE007165	AE007165 Mycobacte
40	118.2	9.9	349606	15	EX842583	EX842583 Mycobacte
41	117.8	9.9	24268	1	AP0070947	AP0070947 Streptomy
42	117.8	9.9	300550	1	AP005030	AP005030 Streptomy
43	117	9.8	8407	1	SEU82823	U82823 Saccharopol
44	116.4	9.8	18109	1	AE006571	AE006571 Mycobacte
45	116.4	9.8	320050	1	BX248336	BX248336 Mycobacte

ALIGNMENTS

RESULT 1
AF015825
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AF015825
Bacillus subtilis 168 cotT-rapA region sequence.
35739 bp DNA linear BCT 07-MAY-1998

GI:2612880

Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 35739)
Rivolta, C., Soldo, B., Lazarevic, V., Joris, B., Muel, C. and Karamata, D.
A 35.7 kb DNA fragment from the Bacillus subtilis chromosome containing a putative 12.3 kb operon involved in hexuronate

Wed


```
/translation="MAGKKTITINGVEMASEEQVLQLLNNSSIEVPOVCHPSLGP
IETDCTCTVINGELKRSASLKDGDVIDTSLSPDKVAQVIGMDKILYNHLYCTVC
DYNNCGCIHTNWKEMKINHOSIPDPKHPYKHDESHPFYRDPDQCLILCGRCEVACQD
VQVETLITDWERKEPRVWDNDVPIINSSCVSCCHSTVPCPNAMKEGMEGAYIL
TGNNETLRPMLEITKGVEYGVSLAIDSMESAMDERLIKTKTCTVCTYCGVGSFV
WTKGRDILKVEPQEPAPANGISTCVKGFQWDFVNSEERLYKPLIRGDHFRABWEE
ALLILASFTLEKAFGPDLSLAFITSSKCTNEESYLMQKLAGVIGTNNVNCRSYQ
SPATAGLFRTVGYGSDGSIIDIAQADLVLIIGNSTSESHFVLSIRIKRAHKLGRQKV
IVADIRKHEMAYEAKETGIDQETLIOMAEIMIGQADSVLCALWAMGVTOHIGSDTSTAIN
LSLAVTGYNKPKGAGSVPLRGHNNVQASDFGSMPLRCPYKVTDEQVROKYRVMVG
PLPKPKNWNEHMIKHSGLKAMVYKGEEMGLVDSNINHVHAAEYKLDFFVQDIF
LSRTAEFADVLVPASLEKGFINTERRIQRLQVFEPIGESKPDQWIIEMEVAANKL
GAGWLYEHPADIMEBAALKSPIYAGTVYERLEGYNLQWPNVADKQSPLLFTFRFPF
PDGKAILYVQWTEPEFCEBYDIHNNRGLLEHFEHGLTYKSGIKSEKTEPEVLEI
SPELAABRGIOQTLVRLTSPFGNVKVKCLIDTRVKGEVLPMDNSGEAAINLLTGS
HADKQDTPAYKETSARKMELIKHDGISPLPKINHNGNFPQIQGVQVHKWARKDYIF
PGDAVKGWGHNG"
7762..8373
/gene="yjgd"
RBS
7762..7769
/gene="yjgd"
7780..8340
/gene="yjgd"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC46313.1"
/db_xref="GI:2612889"
/translation="WAKAKRLOKLEVTEDORKEDELEIDALDHKEAILETHML
CHMNERGVLPRLRGLFGQDGKVLIDLVKKADETEFANTLKNLLIFGLMDLVKQLE
FLILKAGVASAVEKQNSEKTYGFDIIRLSLDKPEINKSLTILFSLKMGQDPTKEL
ERTTQPEHQHQBPREKRGKMKRD"
9347..8373
/gene="yjgd"
RBS
9525..9223
/gene="yjha"
9525..8530
/gene="yjha"
9537..9178
/gene="yjha"
```

Query Match 99.9%; Score 1189.4; DB 1; Length 35739;

```
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAATGTGTTAAACCGCGGCAAGCCTTGCAGCGAGCGCTGCTCAATGGGAAAAACAAA 60
DB 10583 ATGAATGTGTTAAACCGCGGCAAGCCTTGCAGCGAGCGCTGCTCAATGGGAAAAACAAA 10642

QY 61 CAGGATGCGTATCATCCGTTTCCATGGTATGAATCGATGAGAAAGATCGCGCTTTTCC 120
DB 10643 CAGGATGCGTATCATCCGTTTCCATGGTATGAATCGATGAGAAAGATCGCGCTTTTCC 10702

QY 121 TTTGATGAGAAAACCAAGTGTGGAGCGTTTCTTTATGATGATGTCAAAAAGTTGTT 180
DB 10703 TTTGATGAGAAAACCAAGTGTGGAGCGTTTCTTTATGATGATGTCAAAAAGTTGTT 10762

QY 181 GGGGATAAAGAGTGTGTTTCCAGTTTGCATGCGCAGCAGACAGCTCTATTGGAATATCC 240
DB 10763 GGGGATAAAGAGTGTGTTTCCAGTTTGCATGCGCAGCAGACAGCTCTATTGGAATATCC 10822

QY 241 ATCAATTAACATGAGACCGCGGCAAGCATACAAAATCCCGTTTCAGTGTGTAACAAAGCCTTT 300
DB 10823 ATCAATTAACATGAGACCGCGGCAAGCATACAAAATCCCGTTTCAGTGTGTAACAAAGCCTTT 10882

QY 301 ACTCCGCGCGTATGAAGCAATGGGAACCGGAAATTCAGAAATCACAGATGAATGATT 360
DB 10883 ACTCCGCGCGTATGAAGCAATGGGAACCGGAAATTCAGAAATCACAGATGAATGATT 10942

QY 361 CAAAATTTCCGGGGCGAGTGAAGTTGACCTTGTTCAGATTTTTCACCGGCTTCGG 420
DB 10943 CAAAATTTCCGGGGCGAGTGAAGTTGACCTTGTTCAGATTTTTCACCGGCTTCGG 11002
```

```
QY 421 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAGCA 480
DB 11003 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAGCA 11062

QY 481 TGGTCTGATCTTCTGCTCAGTACACCGAAGGATAAAAGTGAAGAGCTGGAAGAGCCTTT 540
DB 11063 TGGTCTGATCTTCTGCTCAGTACACCGAAGGATAAAAGTGAAGAGCTGGAAGAGCCTTT 11122

QY 541 TTGGAAGAACGAGATAACTGTGAGGAAGAACTGGCCCGGTTTTTTTCCGGCATCATAGAA 600
DB 11123 TTGGAAGAACGAGATAACTGTGAGGAAGAACTGGCCCGGTTTTTTTCCGGCATCATAGAA 11182

QY 601 GAAAGCGGAAACAAACCGGAAACAGGATATTATTTCTATTATTAGTGAAGCGGAAAGAAACA 660
DB 11183 GAAAGCGGAAACAAACCGGAAACAGGATATTATTTCTATTATTAGTGAAGCGGAAAGAAACA 11242

QY 661 GCGGAGAAAGCTGTCCGTTGAGAGAGCTGATTCGTTTTTGCACGCTGCTGCTGTGTCGCCGA 720
DB 11243 GCGGAGAAAGCTGTCCGTTGAGAGAGCTGATTCGTTTTTGCACGCTGCTGCTGTGTCGCCGA 11302

QY 721 AATGAACCACTACAAACCTGATTTCAATCGATGTACAGCATATTAGAACGCGCAGC 780
DB 11303 AATGAACCACTACAAACCTGATTTCAATCGATGTACAGCATATTAGAACGCGCAGC 11362

QY 781 GTTTACGAGGAACTGCGCAGCAGCATCTCTGAACCTGATGCTCAGGCAAGTGGAGGAGCCTTTG 840
DB 11363 GTTTACGAGGAACTGCGCAGCAGCATCTCTGAACCTGATGCTCAGGCAAGTGGAGGAGCCTTTG 11422

QY 841 CGTTTCAGAGCGCGCGCGCGCGCTTTTGGAGGCGCATTCGCAAGCGGATACGAGATCGG 900
DB 11423 CGTTTCAGAGCGCGCGCGCGCGCTTTTGGAGGCGCATTCGCAAGCGGATACGAGATCGG 11482

QY 901 GGGCACCTGATTAAAGAAAGTGTATGTTTGGGCTTTGTGGCATCGGCAATCGTGAT 960
DB 11483 GGGCACCTGATTAAAGAAAGTGTATGTTTGGGCTTTGTGGCATCGGCAATCGTGAT 11542

QY 961 GAAGCAAAAGTTTGACAGACCGCACATGTTTGATATCGCCGCCCATCCCAATCCGCATATT 1020
DB 11543 GAAGCAAAAGTTTGACAGACCGCACATGTTTGATATCGCCGCCCATCCCAATCCGCATATT 11602

QY 1021 GCGTTTGGCAGCGGATCCATTTTTCCTTGGGCGCGCGCTTGGCGCTTGTGAAGCAAT 1080
DB 11603 GCGTTTGGCAGCGGATCCATTTTTCCTTGGGCGCGCGCTTGGCGCTTGTGAAGCAAT 11662

QY 1081 ATCGGTTAAAGCTCTTTGATTCTGCTTTTCTCATATGAGTGGTCACTATCACTCG 1140
DB 11663 ATCGGTTAAAGCTCTTTGATTCTGCTTTTCTCATATGAGTGGTCACTATCACTCG 11722

QY 1141 ATTGAACACAGTGTGATATACGATTAAAGAGCTTCCGTTGAAAATGTAA 1191
DB 11723 ATTGAACACAGTGTGATATACGATTAAAGAGCTTCCGTTGAAAATGTAA 11773
```

RESULT 2

```
BSUB0007 201241 bp DNA linear BCT 07-JUL-2003
LOCUS Bacillus subtilis complete genome (section 7 of 21): from 1209742
to 1410982
DEFINITION Z99110 AL009126
ACCESSION Z99110.2 GI:32468738
VERSION Z99110.2
KEYWORDS
SOURCE
ORGANISM Bacillus subtilis subsp. subtilis str. 168
Bacillus subtilis subsp. subtilis str. 168
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 201241)
AUTHORS Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Borrias, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 23:01:12 ; Search time 223 Seconds
(without alignments)
560.292 Million cell updates/sec

Title: US-09-869-334B-1
Perfect score: 2052
Sequence: 1 MVLNRQALQALLGNK.....SITPIENSVIYGLKSRVKN 396

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_podent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rviro.*
- 16: sp_bacterioph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	833.5	40.6	404	Q81Q07	Q81Q07 bacillus an
2	732	35.7	409	Q81CY0	Q81CY0 bacillus ce
3	715	34.8	411	Q81Q12	Q81Q12 bacillus an
4	674	32.8	417	Q59910	Q59910 streptomyce
5	661.5	32.2	402	Q8R221	Q8R221 streptomyce
6	642	31.3	416	Q87605	Q87605 streptomyce
7	634.5	30.9	393	Q82LM3	Q82LM3 streptomyce
8	626	30.5	376	Q81LW5	Q81LW5 bacillus su
9	622	30.3	457	Q82BH2	Q82BH2 streptomyce
10	601	29.3	395	Q8K2M6	Q8K2M6 bacillus su
11	599.5	29.2	388	Q82A10	Q82A10 streptomyce
12	589	28.7	433	Q9F2Q0	Q9F2Q0 streptomyce
13	582.5	28.4	397	Q8KIC8	Q8KIC8 nocardia ae
14	581	28.3	397	Q59523	Q59523 micromonosp
15	581	28.3	434	Q9CBE7	Q9CBE7 mycobacteri
16	576.5	28.1	430	Q8NSW2	Q8NSW2 corynebacte

17	575.5	28.0	406	2	Q83X78	Q83X78 streptomyce
18	571	27.9	418	16	Q9HIR4	Q9HIR4 pseudomonas
19	570.5	27.8	405	16	Q82ES4	Q82ES4 streptomyce
20	566	27.6	401	2	Q9KHJ7	Q9KHJ7 streptomyce
21	565	27.5	399	2	Q8GGQ1	Q8GGQ1 streptomyce
22	564	27.5	404	2	Q54302	Q54302 streptomyce
23	562.5	27.4	413	2	Q848C0	Q848C0 streptomyce
24	561.5	27.4	417	2	Q83WG3	Q83WG3 streptomyce
25	560.5	27.3	433	16	Q89G11	Q89G11 bradyrhizob
26	557.5	27.2	410	2	Q9XSP9	Q9XSP9 streptomyce
27	550	26.8	399	16	Q82QC6	Q82QC6 streptomyce
28	548.5	26.7	404	2	Q83X67	Q83X67 corynebacte
29	547.5	26.7	461	16	Q8FS42	Q8FS42 streptomyce
30	545	26.6	419	2	Q8X332	Q8X332 streptomyce
31	544.5	26.5	407	2	Q59819	Q59819 streptomyce
32	540	26.3	399	16	Q93HA3	Q93HA3 mycobacteri
33	538	26.2	310	2	Q32927	Q32927 mycobacteri
34	536.5	26.1	411	16	Q9X8Q3	Q9X8Q3 streptomyce
35	534.5	26.0	367	17	Q97212	Q97212 sulfolobus
36	533	26.0	396	2	Q9L465	Q9L465 streptomyce
37	532.5	26.0	411	2	Q32460	Q32460 actinomadu
38	532	25.9	404	16	Q93HJ0	Q93HJ0 streptomyce
39	531.5	25.9	418	16	Q89NK6	Q89NK6 bradyrhizob
40	529.5	25.8	417	2	Q8R2J4	Q8R2J4 stigmatella
41	528	25.7	415	16	Q92QQ5	Q92QQ5 rhizobium m
42	527	25.7	396	2	Q9EYL2	Q9EYL2 streptomyce
43	524	25.5	394	16	Q9F2L1	Q9F2L1 streptomyce
44	524	25.5	415	16	Q8UFY8	Q8UFY8 agrobacteri
45	523.5	25.5	416	16	Q982V9	Q982V9 rhizobium l

ALIGNMENTS

RESULT 1

Q81Q07 Q81Q07 PRELIMINARY; 2RT; 404 AA.
AC Q81Q07;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Cytochrome P450 family protein.
GN BAZ632.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beaman M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.B., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT *The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.*;
RL Nature 423:81-86(2003).
DR EMBL; AE017832; AAP26430.1; -
DR TIGR; BAZ632; -
DR GO; GO:0006118; P:electron transport; IEN.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Complete proteome.
SQ SEQUENCE 404 AA; 46728 MW; 960552E1D7792F55 CRC64;

Query Match

40.6%; Score 833.5; DB 16; Length 404;

[illegible]

Qy	66	--FSSCMPQOOTSSIGNS-----IINMDPPKHKTIRSUVNKAFTPRVMKQWEPRIQEITD 117
Db	72	NVFSQDKNMYLSVDNSDHTTTHMLNDEPPNHSRLRSUVQKFTPKMTAQDGRIQRAD 131
Qy	118	ELIQKPGRSEEDLVHDFSPVLPVIVTSELIGVPSAHMEQFKANSDLLVS*PKDKSEAE 177
Db	132	DLISDIERKGTNLVDVDSFPLPIIVISEMLGIDPKEDQAKFRINSHAVIASPETEE--- 188
Qy	178	KAFLEERDKCEEELAAFFAGIIEBKRNKPEODIISILVEAETGKLSGEELIPFCTILL 237
Db	189	--IKETEKQLSEPTITVQLVLDIKRKEPKEDVLVSALILASSEGHKLSARELYSIMILLI 245
Qy	238	VAGNETTNLISNMYKSILETGPGVEELRSHPELMPQAVEALRFPAPV-LRRIAKED 296
Db	246	VAGHETTVNITNTVLALNPNQLKONPKLIDSALIESGLRYSYPVEVTTAAWAEF 305
Qy	297	TEIGGHLIKEGDMVLAFVASANRDEAKFDRPMPDIRBHPNPHIAFGHIGHFCLGAPLAR 356
Db	306	FQIHQTIQGMVWIIALASANDTETVFEIPIDITRENKRIHAFGHSHFCLGAPLAR 365
Qy	357	LEAMIALTSLISAPPHME 374
Db	366	LEAKIATITLFRMPELQ 383
RESULT 3		
Q81Q12		
ID	Q81Q12	PRELIMINARY; PRT; 411 AA.
AC	Q81Q12	
DT	01-JUN-2003	(TrEMBLrel. 24, Created)
DT	01-JUN-2003	(TrEMBLrel. 24, last sequence update)
DT	01-OCT-2003	(TrEMBLrel. 25, last annotation update)
DE	Cytochrome P450.	
GN	CYP4 OR BA2627.	
OS	Bacillus anthracis (strain Ames).	
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
OX	NCBI_TaxID=198094;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=22608414; PubMed=12721629;	
RA	Read T.D., Peterson S.N., Tourasse K., Baillie L.W., Paulsen I.T.,	
RA	Nelson K.E., Tectelin H., Fouts D.E., Eisen J.A., Gill S.R.,	
RA	Holtzapfel E.K., Ostad O.A., Helgason E., Kilstone J., Wu M.,	
RA	Kolony J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,	
RA	DeBoy R.T., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,	
RA	Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,	
RA	Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,	
RA	Berry K.J., Plaut R.D., Wolf A.X., Watkins K.L., Nierman W.C.,	
RA	Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,	
RA	Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,	
RA	Fraser C.M.;	
RT	"The genome sequence of Bacillus anthracis Ames and comparison to	
RT	closely related bacteria.";	
RL	Nature 423:81-86(2003).	
DR	EMBL; AE017032; AAP26475.1; -	
DR	TIGR; BA2627; -	
DR	GO; GO:0006118; P:electron transport; IEA.	
DR	InterPro; IPR001128; Cytochrome_P450.	
DR	Pfam; PF00067; P450; 1.	
DR	PRINTS; PR00385; P450.	
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.	
KW	Complete proteome.	
SQ	SEQUENCE 411 AA; 47050 MW; 3BB2B70A0A8FD1B CRC64;	
Query Match		34.8%; Score 715; DB 16; Length 411;
Best Local Similarity		39.2%; Pred. NO. 8.5e-44;
Matches 148; Conservative 78; Mismatches 126; Indels 26; Gaps		
Qy	14	LLNGKNKODAYHPFHWBESMRKADPVSDENQV---NSVFLYDDVKVVGKEL----- 65
Db	17	LASQFKEDATE---IYKSRKVQVPLFNKTELGAEWLITYEDALPCKNRLKKDPA 73
Qy	66	--FSSCMPQOOTSSIGNS-----IINMDPPKHKTIRSUVNKAFTPRVMKQWEPRIQEITD 117

```

Db 74 NVFSODTLNVELTVDNSDYLTHMLNSOPPNRRLRSIVQVFTPKMAQLGRIQDIAD 133
Qy 118 ELIOKQFQSRFSDLVHDSYPLVIVISLGLVPSAHMEQKAWSDLLVSTPKDSEAE 177
Db 134 DLLNEVERKSGSLNVDVDFPLIIVISEMLGIPKEDQAKFIMSHAWIATPTEEE--- 190
Qy 178 KAFLEERKCEELAAFFAGIIEERKNPEODIISILVEAETSEKLSGEELIPFCTLL 237
Db 191 ---IKETEKQLSEFITYIQLVDMKPKEDVLSALILAESEGHKLSARELYSMIMLLI 247
Qy 238 VAGNETTNLISNAMYSLIETPGVVEERSHPELMPQAVEALRPAPVLRRIAKRD 296
Db 248 VAGHETTNNLTNTVALLLENQQLKENPKLIDAAIEGLRYSPVEVTTSMADPEP 307
Qy 297 TEIGHGLIKEGDWLAFVANSRDRAKDRPMFDIRHPNPHIAFGHICLGLAPLAR 356
Db 308 FOIHQOTIEKGMVWVIALAANRDETTFENPEVFDITRENNRHIAFGHSHFCLGLAPLAR 367
Qy 357 LEANIALTSLISAFPHME 374
Db 368 LEAKIATTLFRMPELQ 385

RESULT 4
ID Q59910 PRELIMINARY; PRT; 417 AA.
AC Q59910;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative cytochrome P450 113B1.
GN CYP113B1.
OS Streptomyces fradiae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T59235;
RX MEDLINE=95075319; PubMed=7984112;
RA Merson-Davies L.A., Cundliffe E.;
RT "Analysis of five tylosin biosynthetic genes from the tyll3A region of
RT the Streptomyces fradiae genome."
RL Mol. Microbiol. 13:349-355(1994).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; U08223; AAA21341.1; -.
DR F1R; S49051; S49051.
DR HSSP; Q55080; I107.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Electron transport; Heme; Membrane;
KW Monooxygenase; Oxidoreductase.
FT METAL 366
SQ SEQUENCE 417 AA; 47224 MW; A761942242PC35B1 CRC64;

Query Match 32.8%; Score 674; DB 2; Length 417;
Best Local Similarity 40.2%; Pred. No. 8.3e-41;
Matches 144; Conservative 64; Mismatches 126; Indels 24; Gaps 4;

Qy 29 WYEMRKDAPVSPDEENQVMSVFLYDDVKVGDKELFSS-----CMPQCTSSI--GNS 80
Db 38 WFAARAEAPVWDESRQAWQVFRYDDYLTVTNPNQVSSDFSPVPVPEELALMGPGT 97
Qy 81 IINMDPKHTKIRSVNKAFTPRVMKQWBPRIQETDTELTKQFGRSEFDLVHDFSYPLP 140
Db 98 FGGIDPPRGPLKLVSAFTPRATLPRIATLGLDLGRKQIDVDSLDVAYPLP 157
Qy 141 VIVISELLGVPSAHMEQKAWSDLLVST-----PKDSEAEAKFALEERKCEELA 192

```

```

Db 158 VVIVASLGLIPABDRDLRFREWVDVITAKNEGMEYPNLPDDFSETMGPAIKEWGDIYFRFA 217
Qy 193 AFFAGIIEBKRNKPDQIISILVEAETSEKLSGEELIPFCTLLVAGNETTNLISNAM 252
Db 218 L-----KRETPDDLSMGLIEAEVEGRRLTDEIVNIVALLTAGHSSATLLGNLF 269
Qy 253 YSILSTPGVVEELRSHPELMPQAVEALRPAPVLRRIAKRDTEIGHGLIKEGDWLFA 312
Db 270 LVLDHREBAQELRADROLIPGAIBETLKYRSPFNFIPELLKEDTDILGHPMKAGQMVVA 329
Qy 313 FVASANRDEAKDRPMFDIRHPNPHIAFGHICLGLAPLARLEANIALTSLISAF 370
Db 330 WIASANRUSAHFSDPTDFVRQPNKHMFGHICLGLFLARLEAKVLELFFDFEF 387

RESULT 5
ID Q8RR21 PRELIMINARY; PRT; 402 AA.
AC Q8RR21;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450 monooxygenase.
GN VISO.
OS Streptomyces virginiae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1961;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21940635; PubMed=11943483;
RA Namwat W., Kamioka Y., Kinoshita H., Yamada Y., Nihira T.;
RT "Characterization of virginiamycin S biosynthetic genes from
RT Streptomyces virginiae."
RL Gene 285:283-290(2002).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB072568; BAB83674.1; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 402 AA; 44632 MW; 622ADF33CFB1C946 CRC64;

Query Match 32.2%; Score 561.5; DB 2; Length 402;
Best Local Similarity 40.1%; Pred. No. 6.4e-40;
Matches 146; Conservative 69; Mismatches 130; Indels 19; Gaps 7;

Qy 29 WYEMRKDAPVSPDEENQVMSVFLYDDVKVGDKELFSSCMPQ---QTSIGNKIN-M 84
Db 22 WLRTWDEHPVHEDYQG-VFHVYHSDVLAVTSDPAVFSDDLRLRFDSSALSEEILSVI 80
Qy 85 DPPKHTKIRSVNKAFTPRVMKQWBPRIQETDTELTKQFGRSEFDLVHDFSYPLVIV 144
Db 81 DPPLHKLRLSVSAFTLRVADLEPRVTELAGRLLEKVEG-SEFOLWGFAYPLVIV 139
Qy 145 SELLGVPSAHMEQKAWSDLLVSTPKDSEAEAKFALEERKCEE-----ELAAFFA 196
Db 140 AELLGVPAEDRELRFQWSDRMLSMQVD--DPLIFQFDEAGEDYERVKEPLKEXHYLQ 197
Qy 197 GIIEERKNPEQDIISILVEAETSEKLSGEELIPFCTLLVAGNETTNLISNAXYSIL 256
Db 198 RHVDARETPGDDLSLVTATIAAGERLTDQIVFEGALLMAGHVSSTMLGKTVLCLE 257
Qy 257 ETGVTVEELRSHPELMPQAVEALRPAPVLRRIAKRDTEIGHGLIKEGDWLFAVAS 316
Db 258 ENDETAALRADRALISGVIEEVLRMRPPIITVAARVTTGEVWVGVTIPKDRMVVASLLS 317
Qy 317 ANDEKAKEPDMFEDIRHPNPHIAFGHICLGLAPLARLEANIALTSLISAFPHMECV 376

```

```

Db 318 ANHDERHIQDVEFDRSRSPNQLAFHGHIYCLGSPGLARLEGRVALEMLLDRF---EDI 374
QY 377 SITP 380
Db 375 RVTP 378

RESULT 6
ID O87605 PRELIMINARY; PRT; 416 AA.
AC O87605;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450 CYP107L1.
GN CYP107L1 OR PICK OR PIKC.
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1936;
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC15439;
RC STRAIN=ATCC15439;
RX MEDLINE=99051447; PubMed=9831532;
RA Xue Y., Wilson D., Zhao L., Liu H.-W., Sherman D.H.;
RT "Hydroxylation of the macrolide P450 hydroxylase from Streptomyces
RL the piKC-encoded cytochrome P450 in Streptomyces venezuelae.";
RN Chem. Biol. 5:661-669 (1998).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC15439;
RC STRAIN=ATCC15439;
RX MEDLINE=98445333; PubMed=9770448;
RA Xue Y., Zhao L., Liu H.-W., Sherman D.H.;
RT "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces
RL venezuelae: architecture of metabolic diversity.";
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF087022; AAC64105.1; -
DR ENBL; AF079139; AAC68886.1; -
DR HSSP; Q00441; IOXA.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; Cytochrome_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 416 AA; 46038 MW; B7392C742045F06B CRC64;

Query Match 31.3%; Score 642; DB 2; Length 416;
Best Local Similarity 37.4%; Pred. No. 1.7e-38;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AHPFPWYENRNDAP---VSFDENQVSVFLYDDVKKVGD---KELFSSCMP- 71
Db 22 QDEAADPYPTVYARLRAGPAHRVTPGDEVLVGVYDRAVLADPRFSKDWENSTTPL 81
QY 72 -QQTSSIGNSTINDDPKHTKIRSVNKAFTPRVMQWEPRIQETIDELQKF---QGR 126
Db 82 TEAEALNNHMLNDDPPRRHTRKLRLKRAEFTRRVELLRVQVIVGLVDAMLAADGR 142
QY 127 SEFDLVHDFSVPLVIVISLLGVPSAHMEQFKAWSDLVSTPKDKSEAEKAFLEERDK 186
Db 142 A--DLMSLAWPLDITVISELLGVPEPDRAAFRWMTDAFVP--PDDPAQAQTAM----- 191
QY 187 CESELAFAFFAGIIEKRNKPEQDIISIVE-ABETGEXLSEELIPFCTLLLVAGNETTT 245

```

```

Db 192 --AEMSGYSLRDLDSKRGQGDGLLSALVTSDEGSRLTSEELGMAHILLVAGHETTV 249
QY STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
Db 246 NLISNAMYSILETPGVYEEELRSHPELMPQAVEALRFAPA-PVLRRIAKEDTEIGGELI 304
Db 250 NLIANGYALLSHPDQLAALRAOWTLTDGAVEMLRYEGVESATYRFPVPEVLDGTVI 309
QY 305 KEGDMLAFVASNRDEAKFDRPHMFDIRHPNPHIAFGHGHICGAPLARLEANTALT 364
Db 310 PAGDTVLVLADAHRTPERFPDRFDIRDTAGHLAFHGHIHFCIGAPLARLEARIAYR 369
QY 365 SLISAPPHMECVSITPIE-----NSVIYGLKSRVK 395
Db 370 ALLERCPDL-ALDVSPGELVWYVNPMPIRGLKALFIR 404

RESULT 7
ID Q82LM3 PRELIMINARY; PRT; 393 AA.
AC Q82LM3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative cytochrome P450.
GN CYP8 OR SAV1987.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RL avermitilis: deducing the ability of producing secondary
RL metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RL microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:528-531 (2003).
DR EMBL; AP005029; BAC69698.1; -
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; Cytochrome_P450; 1.
KW Complete proteome.
SQ SEQUENCE 393 AA; 43126 MW; 1573C19D9A256BA3 CRC64;

Query Match 30.9%; Score 634.5; DB 16; Length 393;
Best Local Similarity 40.2%; Pred. No. 5.7e-38;
Matches 145; Conservative 61; Mismatches 132; Indels 23; Gaps 6;

QY 26 PFPYTESMRKDAPVSF-----DEENQVSVFLYDDVKKVGDKELFSSCYQQTSSI--- 77
Db 18 PYPVYALRERGPVHVVRTPPEAFEGVLVGVHEEAAALADPRLSKDGTKKGLTSLDVE 77
QY 78 --GNSIINMDPKTKTKIRSVNKAFTPRVMQWEPRIQETIDELQKFQGRSEFDLVHDF 135
Db 78 LMGYVLVWDPPEHTRLSLVARAFTRERVALPRIQETIDGLDEMLPRGRADLVDSF 237
QY 136 SYPLPVIVISLLGVPSAHMEQFKAWSDLVSTPKDKSEAEKAFLEERDKCEELAAFF 195
Db 138 APLPVTIVICELGVDPIDIRVTFRLSNIVAPITGGDAELA-----AYERLAAYL 187

```

DR GO:0004497; P:monoxygenase activity; IEA.
DR GO:0006118; P:electron transport; IEA.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450_1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Heme; Hypothetical protein; Yooxygenase; Oxidoreductase;
KW Complete proteome.
SQ SEQUENCE 376 AA; 43418 MW; C3899600FFDB7ECD CRC64;

Query Match 30.5%; Score 626; DB 16; Length 376;
Best Local Similarity 35.4%; Pred.No. 2.2e-37;
Matches 132; Conservative 78; Mismatches 135; Indels 28; Gaps 4

Qy 10 LQRALLGKKNQDAYHPPEFYEMRKDAPV-----SFDEEKQWMSVFLYD----- 54
Db 3 MEKLMFPHGKKEFHNPSPVLGRPREEPPIHRFELKRFGATYPAWLITRYDDCMAFKDN 62

Qy 55 ----DVKKVGVCDKELFSSCMQQTSSIGNSIINMDPPKHTRSVVNKKAFTPRVMKQWEP 110
Db 63 RITRDVKNVWNEQIKMLNVSEIDFVSDHMLAKDTDPHTRLRSLVHQAFTRTIENLRG 122

Qy 111 RIQIETDELIOKFGQRSEFDLWHDPSYLPVTVIVISELIGVPSAHMEQFKAWSDLLVSTPK 170
Db 123 SIEQIAQLIDMEKKNKADIMKSFASPLFVIVISELMGIPKEDRSQFQIWNAMVDTSE 182

Qy 171 DKSEAEKAFLEERDKCEBELAFAFAGIIEKRNKPBDIISILVEAEFTGKLSGEELI 230
Db 183 GNRELITQALREFKD-----YIAKTHDRRIKPKDILSKLVHAEENGSKLSEKELY 234

Qy 231 PFCTLLIVAGNETTNLISNAMYSLIETPGVYEELRSEPELMPQAVEELRFEAPAPVL- 289
Db 235 SMLFLLVWAGLETTVNLGSGTLALQHKKCEKLQKQPEMIATAVEELLRTSPVWMA 294

Qy 290 RRTAKRTEITGGHLIKEGDMVLAFVASANDEAKFDRPHMFDIRHPNPHIAFGHGIFHC 349
Db 295 NRNAIEDFTYKHSIKRGDMIFIGISSNRDPNFENPEILNINSPNPHISFGFGIHF 354

Qy 350 LGAPLARLEANIA 362
Db 355 LGAPLARLEGHIA 367

RESULT 9
ID Q82HH2 PRELIMINARY; PRT; 457 AA.
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative cytochrome P450.
GN CYP16 OR SAV3536.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa T., Hanamoro A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT *Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites".
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa T., Hanamoro A., Shinose M., Kikuchi F., Shiba T.

Sasaki Y., Hattori M., Omura S.;
 "Complete genome sequence and comparative analysis of the industrial
 microorganism Streptomyces avermitilis";
 Nat. Biotechnol. 21:526-531(2003).
 RL EMBL; AP005035; BAC71248.1; --
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; p450; 1.
 KW Complete proteome.
 SQ SEQUENCE 457 AA; 50512 MW; 8342429F7650A635 CRC64;

Query Match 30.3%; Score 622; DB 16; Length 457;
 Best Local Similarity 33.3%; Pred. No. 5.6e-37;
 Matches 137; Conservative 76; Mismatches 148; Indels 50; Gaps 9;

QY 23 AYHPPWYVESMRKQAPV---SPDEENQVMSVFLYDDVKKVWGDKELFSSCMPQQTSSIGN 79
 DB 55 ATDPYFAYAWLREHAPVHRTTLPVGVEAWLVTRYADAKQALADARLSKVPVHSEDA PK 114
 QY 80 S-----IINMDPPKHTKIRSVVNKAFTPRVMQWEPRIQIETDELICKFQ 124
 DB 115 SKTGIPGERSANLTHLMIDPDHTRLRRLVSKAFTPRVAFAFRVQELTDHLIDQPA 174
 QY 125 GRSEDLVHDFYPLPVIVISELLGVPSSAHMEQFKAWSDLLV---STPKDKSEAEKAPL 181
 DB 175 QTSADLIHEFAFPLPIYAICDLLGVPRDQDDFRDWAGMMIRHGGGPRGVARSVK--- 231
 QY 182 EERDKCEEELAAFFAGIIEKRNK-----PEODIISILVEAEETGKLSGEELIPFCTL 235
 DB 232 -----KORGVLAEIIRKREALPADPGPDGLISGLIRASDHGEHLTENEAAMCFV 283
 QY 236 LLVAGNETTNTLISNAMYSIETPGVYBEL-----RSHPELMPQAVEALRFRAPAPVLR 290
 DB 284 LLFAGFETTINLIGNTYALLRNPPQRAKQSIERGEQDLDTGIELLYDGPVELAT 343
 QY 291 -RIAKDDTEIGHLIHEGDMWLAFAVANDDEAKFDRPHMFDIRRHNPNHIAFGHGHFC 349
 DB 344 WRVATEPLDMGGGRIASGDPVILVLAARDPAPFDEPTDLDSRSDNQHLYGHHGHC 403
 QY 350 LGAPLARLANTALTSLSIAFFHMECVSITPIE-----NSVIYGLKSPRYK 395
 DB 404 IGAPLARLESRTALTLRLRLPDLR-LAADPADLRWGGGLMRGLNLPVE 453

RESULT 10
 Q8K2M6 PRELIMINARY; PRT; 395 AA.
 AC Q8K2M6; STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 DT 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE Cytochrome P450 enzyme.
 GN BLOI.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OK2.
 RA Sasaki M., Kotanagi T., Kurusu Y.;
 "Genetic analysis of biotin operon in Bacillus subtilis strain OK2";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AB088066; BAC0324.1; --
 DR GO; GO:0004497; F:monooxygenase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; p450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 395 AA; 44869 MW; 6D4B5C445107D225 CRC64;

Query Match 29.3%; Score 601; DB 2; Length 395;
 Best Local Similarity 34.6%; Pred. No. 1.5e-35;
 Matches 125; Conservative 85; Mismatches 135; Indels -6; Gaps 6;

QY 25 HPPWYVESMRKQAPV---SPDEENQVMSVFLYDDVKKVWGDKEL-----FSSCMPQQT 74
 DB 15 NPYSFYDITLRAVHPHYKGSF-LKYPGWYTYGDETAAILKDAFKVTFPFEESTKYQDL 73
 QY 75 SSIGNS-IINMDPPKHTKIRSVVNKAFTPRVMQWEPRIQIETDELICKFQGRSEFVLV 133
 DB 74 SHVQNMQLFQNPQDHRRLRLTLASGAFTRATESYQPIITETVHQLDQVQGEKQKEVIS 133
 QY 134 DSYPLPVIVISELLGVPSSAHMEQFKAWSDLLVSTPKDKSEAEKAPLEERDKCEEELAA 193
 DB 134 DFAFPLASVFVIANIIGVPEDEQLKWEAASLIQT---IDFTRSRKVLTEGNHMAVOAMA 190
 QY 194 FRAGIITEKRNKPEQDILISILVEAEETGKLSGEELIPFCTLVAGNETTNTLISNAMY 253
 DB 191 YFELLQKRRHPQDMISMLKGE-NDKLTEEAASCTILLAIAGHETTVNLISNVL 249
 QY 254 SILETPGVYEEELASHPELMPQAVEALRFRAPAPVLRRIAKRDTETIGHLIKGDMVLAF 313
 DB 250 CLIQHPEQLLKRENEFDLIGTAVEECLRYESPQMTARVASEDIDISGVTIROGEQVYL 309
 QY 314 VASANDDEAKFDRPHMFDIRRHNPNHIAFGHGHFCGAPLARLANTALTSLSIAFFHME 373
 DB 310 LGAANDRPSIFTNPVDFITRSPNPLSFHGHHVCLGSLARLEAQLAINTFLQRPFL 369
 QY 374 E 374
 DB 370 K 370

RESULT 11
 Q82A10 PRELIMINARY; PRT; 388 AA.
 ID Q82A10; STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 AC Q82A10; MEDLINE=21477403; PubMed=11572948;
 DT 01-JUN-2003 (TReMBLrel. 24, Created)
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Putative cytochrome P450.
 DE CYP23 OR SAV6249.
 GN Streptomyces avermitilis.
 OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sasaki Y., Hattori M.;
 "Genome sequence of an industrial microorganism Streptomyces
 avermitilis: deducing the ability of producing secondary
 metabolites";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinoe M., Kikuchi H., Shiba T.,
 RA Sasaki Y., Hattori M., Omura S.;
 "Complete genome sequence and comparative analysis of the industrial
 microorganism Streptomyces avermitilis";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005046; BAC73960.1; --
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; p450; 1.
 KW Complete proteome.
 SQ SEQUENCE 388 AA; 42468 MW; 1327D228BF068E647 CRC64;

Query Match 29.2%; Score 599.5; DB 16; Length 388;
 Best Local Similarity 34.5%; Pred. No. 1.9e-35;
 Matches 128; Conservative 73; Mismatches 143; Indels 27; Gaps 5;

QY 26 PFPYVESMRKDAV---SFDEENQVMSVFLYDDVKKVGVGDKELFSSCMPQQTSSIGN--- 79
 DB 11 PYAVYDRLEDAPVHRIAGTGGKPAWLWTRYDDVREGIANPLL---SLDKKHALPGNYRG 67

QY 80 -----SIINMDPKHTKIRSVVVKAFTRPVKMQWEPRIQEIITDELIOKFOQRSEPD 130
 DB 68 LALPALDANLLNMDAPDHTIRRLVGRAFTLRRVQLREPVRETAHRLLEDALGTHGSTD 127

QY 131 LVHDFSYPVIVISLGLVPSAHMEQKAMSDLLVSPKQSEBAKAFLEERDKCEE 190
 DB 128 LIASVAAPLPITVICLLGLVGPDEHRDFRAWTDLVTPDAPDPVARESVW-----S 179

QY 191 LAAPFAGIIEKXNKPQEDIIISILVEAETGKLSGEELIPFCTLLVAGNETTINLSN 250
 DB 180 LLGFTGSLADKRNKPADDLLSDLIABOEGDRLTEDELSLAFILLPAGYENTVHLIGN 239

QY 251 AMYSIETPGVYEECSRSHPELMPQAVEALRPAPAPV--LRRIAKRDTIEIGGLIKEGDM 309
 DB 240 AVLALLRHEPQLAA--REDPARLPDVGSEFARVEGFPALLAIREFPVVDVTIGGVTPAGET 299

QY 310 VLAFVASANRDEAKEDRHMEDIRHPNPHIAFGHIHPCIGAPLARLEANALTSLISA 369
 DB 300 VLLSLSAANRDPSPRPDPDLGRDAAGHLALGHGVHYCLGAPLARLETEVALAALLER 359

QY 370 FPHMECVSITP 380
 DB 360 FPDIALAETEP 370

RESULT 12

Q9F2Q0 PRELIMINARY; PRT; 433 AA.
 AC Q9F2Q0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative cytochrome P450 hydroxylase.
 GN SC03099 OR SCE41.08C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1];
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [2];
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [3];
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Kinashi H., Hopwood D.A.;
 RL "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4];
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieger H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT *Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).;
 RL Nature 417:141-147(2002).
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AL939115; CAC09540.1; --
 DR HSSP; C00441; 10XA.
 DR GO; GO:0004497; F-monooxygenase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00607; P450; 1.
 KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.
 SQ SEQUENCE 433 AA; 47351 MW; CDED6D74CA4DA508 CRC64;

Query Match 28.7%; Score 589; DB 16; Length 433;
 Best Local Similarity 32.7%; Pred. No. 1.3e-34;
 Matches 134; Conservative 75; Mismatches 151; Indels 50; Gaps 9;

QY 23 AVHPPEWYESMRKDAV---SFDEENQVMSVFLYDDVKKVGVGDKELF----- 66
 DB 18 ASDPFAVAMLRHAPVHTRLPSSGVAEWLVTRYADAKQALADPRLSKPAHHDEPAHAK 77

QY 67 --SSCMPQQTSSIGNSIINMDPKHTKIRSVVVKAFTRPVKMQWEPRIQEIITDELIOKFO 124
 DB 78 GKTGIPGERKAEALMTHLINIDPDHTLRLRLVSKAFTRPRAVAFPRVQSLADGLDRFA 137

QY 125 GRSEFDLVHDSYPLPVIVISLGLVPSAHMEQKAMSDLLV---STPKDKSEAEKAF 181
 DB 138 DTGSADLIHDFAPLPPIVACLLGLVPRDQDFRDWAGMMIRHQGGPRGVARSYK--- 194

QY 182 EERDKCEELAAFFAGLIEEK-----NKPEQDIISILVEAETGKLSGEELIPFCTL 235
 DB 195 -----KRGYLAHLIRKRAALPEPAPGELLISGLIRADSGHEHLTENEAAAFI 246

QY 236 LLVAGNETTINLSNMYSILETGPVYEELRS-----HPELMPQAVEALRFRAPAPVLR 290
 DB 247 LLFAGFETTVNLWNGTYALLTHPEQRELQTSLAAGERGLLETGVLELLRYDGPVELAT 366

QY 291 -RIAKRDTIEIGGLIKEGDMVLAFVASANRDEAKFDRPHMFDRIRHPNPHIAFGHIHFC 349
 DB 307 WFATRPITIGGQVAAAGDPVLWLAADRDPERFTDPTDLARDSQHLGYGHGHIYC 366

QY 350 LGAPLARLEANALTSLISAFPHMECVSITPIE-----NSVIYGLKSPRV 394
 DB 367 LGAPLARLEGQTALATLLTLRLFDLR--LAADPAELRWGGLIRGLTLPLV 415

RESULT 13

Q8KIC8 PRELIMINARY; PRT; 397 AA.
 AC Q8KIC8;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Cytochrome P450 enzyme (Putative cytochrome P450) (Putative P450 protein) (Cytochrome P-450 Rebd).
 GN RBME OR REBT.
 OS Nocardia acrocolonigenes.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Pseudonocardiaceae; Actinosynnemataceae; Lechevalieria.
 OX NCBI_TaxID=68170;
 RN [1];
 RP SEQUENCE FROM N.A.
 RA Hyun C.-G., Billiign T., Liao J., Thorson J.S.;
 RT "The Biosynthesis of Indolocarbazoles in a Heterologous E. coli Host."
 RL Submitted (AJG-2002) to the EMBL/GenBank/DBJ databases.
 RN [2];
 RP SEQUENCE FROM N.A.

RA Nishizawa T., Sherman D.H.;
 RT "Identification and analysis of the rebeccamycin biosynthetic gene
 RT cluster in *Lechevalieria aerocolonigenes* ATCC39243.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 39243;
 RX MEDLINE=21979046; PubMed=11983340;
 RA Sanchez C., Brana A.F., Mendez C., Salas J.A.;
 RT "Rebeccamycin biosynthesis gene cluster.";
 RL Chem. Biol. 9:519-531 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Onaka H., Taniguchi S., Igarashi Y., Furumai T.;
 RT "Cloning of the biosynthetic gene cluster of rebeccamycin, an
 RT indolocarbazole antitumor agent, from *Lechevalieria aerocolonigenes*
 RT ATCC39243.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AF534707; AA01211.1; -.
 DR EMBL; AB090952; BAC10677.1; -.
 DR EMBL; AJ414559; CAC93717.1; -.
 DR EMBL; AB071405; BAC15753.1; -.
 DR GO; GO:0004497; F:monooxygenase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; p450; 1.
 DR PRINTS; PR00385; p450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 397 AA; 43092 MW; 0286C67FA118731F CRC64;
 SQ

Query Match 28.4%; Score 582.5; DB 2; Length 397;
 Best Local Similarity 37.2%; Pred. No. 3.4e-34;
 Matches 139; Conservative 60; Mismatches 134; Indels 41; Gaps 9;
 QY 22 DAYHPPFWESMRKDPVSDDEENQVMSVFLYDDVKVWGDKL-----FSSCMPQQT 74
 DB 13 ELADPVEVREYLTGDFVHNGE--AMVYFSYDGVAVHLTSRDYGRGPGGRATPIPPSH 70
 QY 75 SSGT-----NSLIMNDPKTKIRSVVNKAFTRPVKMQWEPRIQEIITDELQKQGRSEFD 130
 DB 71 DILSRIVENWLVLDPPRTHTLSLLAKERSPAVVTGLRERVKXIAGELLAGLDAGEID 130
 QY 131 LVHDFSYPLPVIVISELGVPSAHMEQFAW-----SDLLVSTPKDSEAEKA----- 179
 DB 131 LVDFFAAPLPILVISELGVPA-----RLRSWFRRCVLDL-----QEASTARATNP 177
 QY 180 -FLEERDKCEEELAAEFAGIIIEKKNKPEQDIISILVEAEETGKLSGSELIPFCITLL 237
 DB 178 GALARADGAASSELVEFFGG--ELGTRKPDDELDVALLVNAQRGEALTDEIVSTCVHLL 235
 QY 238 VAGNETTTNLISNAMYISILETPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDT 297
 DB 236 TAGHEITTNLISKVALLANPAAAFPLAGLDVTPQVVEELNRFDPVQVYVFWAHQDT 295
 QY 298 EUGHLIKEGDMVLAFFVANSANROGAKFDRPHMFDIRHPNPHIATFGHIFCLGAPLARL 357
 DB 296 ALGGKPIRRGKVVVLGSANROPAAFAEPDRDLDRDSRRHCGFGIGIHYCLGAALART 355
 QY 358 EANTALTSLSIAFP 371
 DB 356 EAEIGLSVLFTNFP 369

RESULT 14
 Q59523 PRELIMINARY; PRT; 397 AA.
 AC Q59523;
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Cytochrome P450 107E1 (Mycimycin biosynthesis protein mycG).

GN CYP107E1 OR MYCG.
 OS Micromonospora griseorubida.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micromonosporineae; Micromonosporaceae; Micromonospora.
 OX NCBI_TaxID=28040;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=All1725;
 RX MEDLINE=95107242; PubMed=7808395;
 RA Inouye M., Takada Y., Muto N., Horinouchi S., Beppu T.;
 RT "Characterization and expression of a p-450-like mycinamicin
 RT biosynthesis gene using a novel Micromonospora-Escherichia coli
 RT shuttle cosmid vector.";
 RL Mol. Gen. Genet. 245:456-464 (1994).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; D16098; BAA03672.1; -.
 DR PIR; S51594; S51594.
 DR HSPF; Q00441; I0XA.
 DR GO; GO:0004497; F:monooxygenase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; p450; 2.
 DR PRINTS; PR00385; p450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 397 AA; 44331 MW; 0FE6A251BESSE233 CRC64;
 SQ

Query Match 28.3%; Score 581; DB 2; Length 397;
 Best Local Similarity 37.3%; Pred. No. 4.4e-34;
 Matches 134; Conservative 64; Mismatches 131; Indels 30; Gaps 8;
 QY 30 YESMRKDPVSDDEENQVMSVFLYDDVKVWGDKLFFSSCMPQQT-----SS 76
 DB 23 YGSELQETEPVSRVPPYGE--AWLVTRYEDVRAVLGDGRFVRG--PSMTRDEPTREPM 78
 QY 77 IGNSIMNDPKTKIRSVVNKAFTRPVKMQWEPRIQEIITDELQKQGRSE-FDLVHDF 135
 DB 79 VKGLLSMDPEHSRLRLVVKAFTRPRAESLRPRASEIAHVLVDQMAATQPADLVAMP 138
 QY 136 SYPLPVTVISELGVPSAHMEQFAWNSCLIVSTPKDSEAEKAFLEERDKCEEELAAFF 195
 DB 139 ARQLPVAVICELLGVPSADHDFRMSGAFSLTAETAEEMQEA-----AEQAYAYM 190
 QY 196 AGTIEKKNKPEQDIISILVEAEETGKLSGSELIPFCITLLVAGNETTNLISNAMYI 255
 DB 191 GDLIDRRKKEPTDLDVLSALVQARDQDLSQEQLLDLAIGLLVAGYESTTQIADFWLL 250
 QY 256 LETPGVYELRSHPELMPQAVEALRFRAPAPV---LRIAKRDTIEGHLIKEGDMVLA 312
 DB 251 MTRPELRQLLDRELIPSAVEELTRW-VPLGVGTAFPRYAVEDVTLRGVTRAGEPVLA 309
 QY 313 FVASANDEAKFDRPHMFDIRHPNPHIATFGHIFCLGAPLARLEANIALTSLSIAFP 371
 DB 310 STGAANRDOAQFPDADRIDVDRTFNQHLGFGHVGHCIGAPLARVELQVALEVLQRLP 368

RESULT 15
 Q9CBE7 PRELIMINARY; PRT; 434 AA.
 AC Q9CBE7;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Putative cytochrome p450.
 GN M12088.
 OS Micobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW;
 RX MEDLINE=21128732; PubMed=11234002;
 QY

RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream K.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 403:1007-1011(2001).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AL583924; CAC31043.1; -.
DR PIR; C87170; C87170.
DR HSSP; Q00441; LJJP.
DR Leproma; ML2088; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ SEQUENCE 434 AA; 47154 MW; 60A83B14B2769067 CRC64;

Query Match 28.3%; Score 581; DB 16; Length 434;
Best Local Similarity 35.1%; Pred. No. 5e-34;
Matches 136; Conservative 80; Mismatches 140; Indels 32; Gaps 6;

QY 8 QALQFALLNGKKNQDAYHPFMYESMRKDAVSF-DEENQVWSVLYDD--VKVVVGDKK 64
DB 35 QVLLLELLDPTGRAD---PFPYRALIDYGPMLTVPFSSDCDEALRHPLSASD 91
QY 65 LFSQCMPOQTSG-----NSIINDPPKHTKIRSVVNKAFTPRVMKQWEPRIQET 116
DB 92 RLKATLAQAATAAGAEPRPFYASSPFMELDPDPTRLRLKLVSKAFAPKVQALEGDIALLV 151
QY 117 DELIQFGQRSEFDLVHDFSYPLPVIVISELGVPSAHMEQKAWSDLLVST----- 168
DB 152 DSLIDKGAAGQFDVIADLAPLAVAVICRLIGVPVEDAPEFGVRSALLVQSVDPFITIT 211
QY 169 --PKDKSEAEKAFLEERDKCEEELAAFPAGLIEKKNKPEODIISILVEAETGKLSG 226
DB 212 GEPPTEATERLRAGVLRDYLEQ-----LVKCRGTGEGEDLISRLIEDESQOQLTE 263
QY 227 EELIPFCCLLVAGMETTNTNISNAMYSLIETPGVYSRLSHPLMPQAVEALFRAPA 286
DB 264 EEIATCGLLLVAGHTTWNLIANAVLMLRNPSQWKALSSNPQAPLVVEETLRYDPAT 323
QY 287 PVLRIAKRDETEIGGHLIKEGDMVLAFVASANRBAKDRPHMFDIRRHNPPIAFGHCI 346
DB 324 HLIQRVAARKMTIGTTLTEGDTWLLLAANRDPAVYSRDPDFDPDRPSSRHAFVAGS 383
QY 347 HFCIGAPLARLENNIALTSLISAPFME 374
DB 384 HFCIGALARLEATVTLISAIRFFQVQ 411

Search completed: May 28, 2004, 23:22:10

Job time : 225 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 23:22:17 ; Search time 352 Seconds
(without alignments)
314.156 Million cell updates/sec

Title: US-09-869-334B-1

Perfect score: 2052

Sequence: 1 MNVLRNRQALQRLANGK.....SITPIENSVIYGLKSRVKM 396

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0

Maximum DB seq length: 2003000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap:
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap:
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap:
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap:
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap:
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap:
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pap:
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pap:
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pap:
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap:
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap:
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap:
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pap:
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap:
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap:
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	670	32.7	410	14	US-10-205-032-6
2	649.5	31.7	404	14	US-10-214-446-40
3	642	31.3	416	9	US-09-861-289-39
4	642	31.3	416	9	US-09-860-846-39
5	642	31.3	416	10	US-09-988-384B-39
6	642	31.3	416	10	US-09-836-821-39
7	642	31.3	416	10	US-09-793-708-18
8	642	31.3	416	14	US-10-201-365-13
9	642	31.3	416	14	US-10-160-539-18
10	642	31.3	416	14	US-10-271-889-39
11	636	31.0	425	14	US-10-214-446-20
12	634.5	30.9	393	14	US-10-156-761-9525
13	623.5	30.4	392	14	US-10-214-446-32
14	622	30.3	457	14	US-10-156-761-11073
15	620.5	30.2	403	16	US-10-229-148B-20

16	599.5	29.2	388	14	US-10-156-761-13776
17	587.5	28.6	408	14	US-10-214-446-4
18	586	28.6	404	14	US-10-214-446-50
19	576.5	28.1	430	9	US-09-738-626-4117
20	571	27.8	418	12	US-10-389-647-559
21	570.5	27.8	405	14	US-10-156-761-12073
22	565	27.5	399	14	US-10-314-657-37
23	557	27.1	404	14	US-10-321-188-76
24	550	26.8	399	14	US-10-156-761-8126
25	550	26.8	432	14	US-10-145-415-10
26	549	26.8	429	14	US-10-145-415-14
27	549	26.8	430	14	US-10-145-415-30
28	549	26.8	430	14	US-10-145-415-95
29	548	26.7	418	16	US-10-461-194-118
30	547	26.7	430	14	US-10-145-415-8
31	547	26.7	430	14	US-10-145-415-20
32	544	26.5	430	14	US-10-145-415-12
33	540	26.3	399	14	US-10-156-761-3914
34	540	26.3	430	14	US-10-145-415-26
35	539	26.3	415	14	US-10-214-446-46
36	537.5	26.2	430	14	US-10-145-415-4
37	537.5	26.2	430	14	US-10-145-415-28
38	536	26.1	430	14	US-10-145-415-18
39	532	25.9	404	14	US-10-156-761-10431
40	531	25.9	430	14	US-10-145-415-24
41	531	25.9	430	14	US-10-145-415-32
42	530	25.8	426	14	US-10-145-415-34
43	530	25.8	430	14	US-10-145-415-2
44	530	25.8	475	14	US-10-145-415-22
45	528	25.7	415	16	US-10-461-134-88

ALIGNMENTS

RESULT 1

US-10-205-032-6

; Sequence 6, Application US/10205032

; Publication NO. US20030113874A1

; GENERAL INFORMATION:

; APPLICANT: Farnet, Chris

; APPLICANT: Yang, Xianshu

; APPLICANT: Staffa, Alfredo

; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARMINICIN

; FILE REFERENCE: 3016-205

; CURRENT APPLICATION NUMBER: US/10/205,032

; CURRENT FILING DATE: 2002-07-26

; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 6

; LENGTH: 410

; TYPE: PRT

; ORGANISM: micromonospora carbonacea subspecies aurantiaca

US-10-205-032-6

Query Match

Best Local Similarity 32.7%; Score 670; DB 14; Length 410;

Mismatches 143; Conservative 65; Mismatches 141; Indels 8; Gaps 2;

QY 22 DAYHFPWVESRKADAPUSFDENQVSWFLDDVKKVGVGDKELFSS-----CMPCQTS 75

DB 19 DAQGLDMFAYNRKKNVPSWSDTRQAWHVSFVSDYQVTVTNLFSSDFTSVFFVSELA 78

QY 76 SI--GNSIINMPPKHRTKIRSVVNKAFTRVMKQWEPRIQETDELQFQGRSEGLVH 133

DB 79 LLMGPTGGIDPPHAPLRKLVSAQTPRRAQWELRGQITADVLDQVROQDRDIAS 138

QY 134 DFSYLPVIVISGLKVSFAHMEQFKANSDDLIVSTPKKSEFAEKAFLEERKCEBELAA 193

DB 139 DLAYLPVTVIAELLGIPTKDHEKFEWDIILSNEGLEYVNPDPDFTETVGPABEWS 198

QY 194 FRAGTIEEKRNKPEODIISILVEASTGKLSGELIPCTILLVAGNTTNTLNISNMY 253

Db 199 FCYQIAHKAABPKODLISGLCAAAYDGRKLTDEEVNIVALLTAGHISSATLLSNLFL 258
QY 254 SILETPGVYEEIRSHPELPOAEALRFRAPAVLRIAKRDEITGHLIKEGDMVLAF 313
Db 259 VLEHFPQAAVRAORSUVGVIEETLRVSPNCIFRILNEDTDILGHPMRKGQWIAW 318
QY 314 VASANDEAKFDRPHMFDIRRHNPPIAFHGHIHFCGLGAPLARLEANIATLSISAF 370
Db 319 IASANKDTEVFTDPTDFDIRRESNKHAFHGHIHFCGLGAPLARLEAKVFNQTLDOF 375
RESULT 2
US-10-214-446-40
; Sequence 40, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-40
Query Match 31.7%; Score 649.5; DB 14; Length 404;
Best Local Similarity 38.3%; Pred. No. 2.1e-49;
Matches 143; Conservative 69; Mismatches 122; Indels 39; Gaps 8;
QY 26 PPFWEKMRKAPVSPDENQ-----VWSVFLYDDVKVVGDKELRSC-- 69
Db 18 PPFNEADGISLADAYEAREQGLLRVMAYGEPAWLATRYADARLVLDGRR-FSRAEG 76
QY 70 ----MPQOTSIGNS-IINMPPKHTKIRSVVNKAFTPRVMKQWEPRIQEIITDELIOKFQ 124
Db 77 ARHDEPRQSEGRDRSDGILSMQPDHTRTLRTLVAKAFTMHQVEKLRPAVRELADLIDKMV 136
QY 125 GR-SEFDLVHDSYPLVPVIVISELLGVPSAHMEQFKAWSDDLIVSTPKKSEAEAKFLEE 183
Db 137 ATGAPVDLVVEEFAFPVGVVICQLLGVPEVDRPRFRAWSDAALST-----SSLTABE 188
QY 184 ROKCEELAAFTAGIIEEKRNKEQDIISILVEABETGKLSGELIIPFCTLLLVAGNET 243
Db 189 FRANGEEELRAYRGLIEDHRAPREDLITGLTEARPDRLTEQELVLDVLCVGLVAGHET 248
QY 244 TTNLSNAMYSILETPGVYEEIRSHPELPOAEALRFRAPAV-----LRIARDETE 298
Db 249 TATQIPNFVVTLDRPEQWNLREDELPVTAVEELMRF---VFLSGASFPYATVEDVE 305
QY 299 IGHILKEGDMVLAFVASANDEAKFDRPHMFDIRRHNPPIAFHGHIHFCGLGAPLARLE 358
Db 306 VGGTLVRAGEPVLVAVGAANRDPAREDAQCQLDLAREGNQHLGFGHGVHFCGLGAPLARLE 365
QY 359 ANIATLSISAPP 371
Db 366 LQELCALGALRLP 378

RESULT 3

US-09-861-289-39

; Sequence 39, Application US/09861289

; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-39
Query Match 31.3%; Score 642; DB 9; Length 416;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;
QY 21 QD-AYHPPFWYEMRKDAP---VSEDEENQVNSVFLYDDVKVVGDKELIOKF-- 71
Db 22 QDFAADPYTYARAEAGFAHRVTRTEGDEWLVGVYDARAVLADPRFSKDMWNSSTPL 81
QY 72 -QOTSIGNSIINMPPKHTKIRSVVNKAFTPRVMKQWEPRIQEIITDELIOKF-- 126
Db 82 TEAEALNHNWLESDDPRHTRILKIVAREFTMRVRELLRFRVQEIIVDGLVDAMLAAPDGR 141
QY 127 SEFDLVHDSYPLVPVIVISELLGVPSAHMEQFKAWSDDLIVSTPKKSEAEAKFLEE 186
Db 142 A--DMESLAWP-LITVISELLGVPEPDRAAPRWMDAFVF--PDDPAQAQTAM----- 191
QY 187 CEEELAAFTAGIIEEKRNKEQDIISILVE-ABETGKLSGELIIPFCTLLLVAGNETTT 245
Db 192 --AEMSGYLSRLIDSKRGQDGEDLLSALVRTSDGSRITSELLGMAHILLVAGHETT 249
QY 246 NLISNAMYSILETPGVYEEIRSHPELPOAEALRFRAPAVLRIAKRDETEIGHLLI 304
Db 250 NLIANQMYALLSHPDQIALARADMTLLDGAVEEMKLYEGEVESATYRFFVPEVDLGTVI 309
QY 305 KEGDMVLAFVASANDEAKFDRPHMFDIRRHNPPIAFHGHIHFCGLGAPLARLEANIAT 364
Db 310 PAGDTVLVVLADAHRTPERFPDPRDIRRDTAGHLAFHGHIHFCIGAPLARLEARI 369
QY 365 SLISAFPHMECVSITPIE-----NSVIYGLKSEFVK 395
Db 370 ALLERCPDL-ALCVSPGELVWYFNPMPMIRGLKALPIR 404
RESULT 4
US-09-860-846-39
; Sequence 39, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT


```
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatenLin Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-10-160-539-18

Query Match      31.3%; Score 642; DB 14; Length 416;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AVHPFPWYEMRKDAP---VSFDEENQVSVFLYDDVKVVGDP---KELFSSCMP- 71
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 72 -QOTSSIGNSIINMDPPKHTKIRSVVKNKAFTRPVKQMPRIQEIITDELIQKF---QGR 126
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQFKAWSDDLVTSTPKKSEAEKAFLEERDK 186
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 142 A--DLMESLAWPLPIITVISELLGVPEPDRAAFRVWTDVVF--PDDPAQAQTAM----- 191
QY 187 CEEELAAFPAGIIEBKRNKPEQDIISILVE-AEETGEKLSGEBELIFPCTLLVAGNETTT 245
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 192 --AEMSGYLSRLIDSKRGQDGEDLLSALVRTSDEGSRLTSEELGMHILLVAGHETTIV 249
QY 246 NLISNAMYSLILETPGVYELRSHPELMPQAVEALRFRAPA-PVLRRIAKRDTETIGGHLI 304
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 250 NLIANGMYALLSHPDQALALRADMTLLDGAVEMLRYEGVESATYRFPVPEVDLGTVI 309
QY 305 KEGDWLAFVASANRDEAKFDRPHMFDIRRHNPHTAFGHGHIFCLGAPLARLEANIALT 364
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 310 PAGDTVLVLADAHRTTPERPDPHREDIRRDRTAGH:AFGHGHIFCGIGAPLARLEARI 369
QY 365 SLISAPFPMCEVSIPTIE-----NSVIYGLKSRVVK 395
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 11
US-10-214-446-20
; Sequence 20, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 6C/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-20

Query Match      31.0%; Score 636; DB 14; Length 425;
Best Local Similarity 33.8%; Pred. No. 3.7e-48;
Matches 137; Conservative 79; Mismatches 145; Indels 44; Gaps 8;

QY 23 AVHPEPWYEMRKDAPV---SFDEENQVSVFLYDDVKVVGDKELF----- 56
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 25 ASDPYAYANLREHAPVHRTTLPSGVSEAWLVTRYGDARQALADQRLSKNPAHHSFHAK 34
QY 67 --SSCMPQOTSSIGNSIINMDPPKHTKIRSVVKNKAFTRPVKQMPRIQEIITDELIQKF 124
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 85 GKTGIPGERKAEMLTHLINIDPPDHTLRRLRLVSKAFTPRRVAETFRVQELTDRIDAFV 144
QY 125 GRSEFDLVHDFSYPPLVIVISELLGVPSAHMEQFKAWSDDLVTSTPKKSEAEKAFLE 181

; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatenLin Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-10-160-539-18

Query Match      31.3%; Score 642; DB 14; Length 416;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AVHPFPWYEMRKDAP---VSFDEENQVSVFLYDDVKVVGDP---KELFSSCMP- 71
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 72 -QOTSSIGNSIINMDPPKHTKIRSVVKNKAFTRPVKQMPRIQEIITDELIQKF---QGR 126
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQFKAWSDDLVTSTPKKSEAEKAFLEERDK 186
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 142 A--DLMESLAWPLPIITVISELLGVPEPDRAAFRVWTDVVF--PDDPAQAQTAM----- 191
QY 187 CEEELAAFPAGIIEBKRNKPEQDIISILVE-AEETGEKLSGEBELIFPCTLLVAGNETTT 245
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 192 --AEMSGYLSRLIDSKRGQDGEDLLSALVRTSDEGSRLTSEELGMHILLVAGHETTIV 249
QY 246 NLISNAMYSLILETPGVYELRSHPELMPQAVEALRFRAPA-PVLRRIAKRDTETIGGHLI 304
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 250 NLIANGMYALLSHPDQALALRADMTLLDGAVEMLRYEGVESATYRFPVPEVDLGTVI 309
QY 305 KEGDWLAFVASANRDEAKFDRPHMFDIRRHNPHTAFGHGHIFCLGAPLARLEANIALT 364
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 310 PAGDTVLVLADAHRTTPERPDPHREDIRRDRTAGH:AFGHGHIFCGIGAPLARLEARI 369
QY 365 SLISAPFPMCEVSIPTIE-----NSVIYGLKSRVVK 395
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 10
US-10-271-889-39
; Sequence 39, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600.582US1
; CURRENT APPLICATION NUMBER: US/10/271,889
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-10-271-889-39
```



```

; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11073
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11073

Query Match      30.3%; Score 622; DB 14; Length 457;
Best Local Similarity 33.3%; Pred. No. 7.4e-47;
Matches 137; Conservative 76; Mismatches 148; Indels 50; Gaps 9;

QY 23 AVHPFWMYEMRKADPV-----SPDEENQVMSVFLYDDVKVVGDKELFSSCMPQQTSSIGN 79
Db 55 ATDPYPAYAWLREHAPVHRTTLPSPGVEAWLVTRYADAKQALADARLSKNPVHSHSEDAPGK 114
QY 80 S-----IINMDPKHTKIRSVNKAFTPRVMKQWEPRIOGITDELICKFCQ 124
Db 115 SGTGIPGERSANMLTHLLNIDPPDFTLRRLRSKAFTRPRVAEFAPRVQELTDHLIDQFA 174
QY 125 GRSEFDLVHDFSYPLVPIVISSELLGVPSAHMEOFKAWSDLLV---STPKDKSBEAKAFL 181
Db 175 QTGSALILHEFAPPLPIYALCDLGVPREQDDFDWAGMMIRHGGGPRGCVARSVK--- 231
QY 182 BERCKSEELAAFFAGLIEEKKNK-----PEQDIISILVEABETCKLSGEBLIPFCIL 235
Db 232 -----KMRGYLABLIHRKREALPADGPGEDLISGLIRASDHGEHLTENEAAAMCFV 283
QY 236 LTVAGNETTNNLISNAMYSILETGPGVVEEL-----RSHPELMPQAVEEALRPRAPAPVLR 290
Db 284 LTFAGFETIINLIGNTYALLRNPPQARLQASIERGEQDLDTGTSEELRYDGPVELAT 343
QY 291 -RIAKRDEIGGHLIKEGDMVLAFVASANRDEAKPDRPHMFDIRRHNPPIAFHGHIFC 349
Db 344 WRYATEPLDMGGQRIASGDPVLVWLAADRDPAFDEPDTLDSRDNQHLGYGHGHIYC 403
QY 350 LGAPLARLEANTLSLSAPPHMECVSTIPTE-----NSVIYGLKSEPRUK 395
Db 404 IGAPLARLEGRATLTLRLRLPOLR--EADPADLRWRGGLIMRGLRLENPEVE 453

```

RESULT 15

```

US-10-229-148B-20
; Sequence 20, Application US/10229148B
; Publication NO. US20040091975A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midecamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/10/229,148B
; PRIOR FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Streptomyces mycarofaciens
US-10-229-148B-20

```

```

Query Match      30.2%; Score 620.5; DB 16; Length 403;
Best Local Similarity 40.2%; Pred. No. 8.4e-47;
Matches 140; Conservative 55; Mismatches 144; Indels 9; Gaps 3;

QY 29 WYESMRKQAPVSPDEENQVMSVFLYDDVKVVGDKELFSS-----CMPQQTSSIGNSI 81
Db 23 WFAFNTRTHHEVFWDSESHAWQVFRYDDLYLTVSNNPQFFSSDFNEVMTTPPELEWVIQGT 82
QY 82 IN-MDPPKHTKIRSVNKAFTPRVMKQWEPRIOGITDELICKFCQGRSEFDLVHDFSYPLP 140
Db 83 IGALDPPAHGPMKLVSAFTPRMARLEPRIRAVIQELLDVAVRGQETIDVVGDLXYALP 142
QY 141 VIVISELLGVPSAHMEOFKAWSDLLVSTPKDKSEEAKEAFLEERDKCEELAAFFAGLIE 200
Db 143 VIVIAELLGIPSGDRDVFRCGWDVTLTNEGLEYPNLPDNFSETTAPALKEMTDYLLHQIH 202
QY 201 EKENKPEQDIISILVEAEETGKLSGEBLIPFCITLLAVAGNETTNNLISNAMYSILETGP 260
Db 203 AKREAPVDDLLISGLVQAEQGRKLTJVEIVNIIVALLTAGHVSSTLSLNLFLVLEENPQ 262
QY 261 VYBELRSHPELMPQAVEEALRPRAPAPVLRRIARRDTEIGGHLIKEGDMVLAFVASANRD 320
Db 263 ALADLRADRELVTGAVEETLRYSPFNNIIFRLKEDTDILGPEMKXQVVIAMWSQSANRD 322
QY 321 EAKPDRPHMFDIRRHNP-HIAFGHGHIFCGLGAPLARLEANTLSL 367
Db 323 PEHFPSPDTFDIRRSSSRHMAFGIGIHHLCLGAFARQEGKVVLLELXL 370

```

Search completed: May 28, 2004, 23:54:59
Job time : 353 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 23:12:17 ; Search time 92 Seconds
(without alignments)
222.216 Million cell updates/sec

Title: US-09-869-334B-1

Perfect score: 2052
Sequence: 1 MNVLRQALQALLGNK.....SITPIENSVIYGLKSRVYM 396

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgm2_6/ptodata/2/iaa/5A COMB.pep.*

2: /cgm2_6/ptodata/2/iaa/5B COMB.pep.*

3: /cgm2_6/ptodata/2/iaa/6A COMB.pep.*

4: /cgm2_6/ptodata/2/iaa/6B COMB.pep.*

5: /cgm2_6/ptodata/2/iaa/PCTUS COMB.pep.*

6: /cgm2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	642	31.3	416	3	US-09-320-878-18
2	642	31.3	416	3	US-09-105-537-39
3	642	31.3	416	4	US-09-141-908-13
4	642	31.3	416	4	US-09-657-440-18
5	606.5	29.6	399	4	US-08-765-907A-10
6	578	28.2	437	4	US-09-252-991A-17836
7	518.5	25.3	419	3	US-09-335-403-8
8	518.5	25.3	419	3	US-09-413-814-71
9	518.5	25.3	419	4	US-09-568-102-8
10	518.5	25.3	419	4	US-09-567-963-8
11	518.5	25.3	419	4	US-09-568-480-8
12	518.5	25.3	419	4	US-09-568-486-8
13	518.5	25.3	419	4	US-09-568-472-8
14	518.5	25.3	419	4	US-09-567-899-8
15	475	23.1	406	6	5212296-6
16	462	22.5	403	6	5212296-9
17	435.5	21.2	409	3	US-09-385-028-12
18	435.5	21.2	409	4	US-09-726-614-12
19	435.5	21.2	409	4	US-09-385-040-12
20	424	20.7	412	1	US-08-102-863-11
21	424	20.7	412	5	PCT-US92-10885-11
22	406.5	19.8	468	4	US-09-252-991A-32437
23	379	18.5	395	4	US-09-266-965-129
24	306	14.9	189	4	US-09-679-279-20
25	263	12.8	443	2	US-09-096-982-9
26	263	12.8	443	2	US-08-653-650A-9
27	262	12.8	422	2	US-09-096-982-5

28	262	12.8	422	2	US-08-653-650A-5	Sequence 5, Appli
29	262	12.8	474	2	US-09-096-982-8	Sequence 8, Appli
30	262	12.8	474	2	US-08-653-650A-8	Sequence 8, Appli
31	258	12.6	422	1	US-08-396-218-2	Sequence 2, Appli
32	258	12.6	422	1	US-08-760-116-2	Sequence 2, Appli
33	246	12.0	444	3	US-09-413-814-95	Sequence 95, Appli
34	246	12.0	444	3	US-09-413-814-96	Sequence 96, Appli
35	213.5	10.4	507	1	US-08-457-274A-23	Sequence 23, Appli
36	213.5	10.4	507	5	PCT-US95-05758-23	Sequence 23, Appli
37	209	10.2	503	3	US-09-144-367-2	Sequence 2, Appli
38	207.5	10.1	382	3	US-09-320-878-7	Sequence 7, Appli
39	207.5	10.1	382	4	US-09-141-908-7	Sequence 7, Appli
40	207.5	10.1	382	4	US-09-657-440-7	Sequence 7, Appli
41	207.5	10.1	402	3	US-09-105-537-22	Sequence 22, Appli
42	207.5	10.1	3782	3	US-09-105-537-4	Sequence 4, Appli
43	207	10.1	501	3	US-09-111-730-1	Sequence 1, Appli
44	200.5	9.8	467	4	US-09-126-420A-17	Sequence 17, Appli
45	199	9.7	501	3	US-08-906-791-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-320-878-18
; Sequence 18, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT CARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-18

Query Match	31.3%;	Score 642;	DB 3;	Length 416;
Best Local Similarity	37.4%;	Pred. No. 2.4e-54;		
Matches 148;	Conservative 77;	Mismatches 137;	Indels 34;	Gaps 12;
QY	21	QD-AVHPFWYEMRKDA2---	VSEDEENQVMSVFLYDDVKXVGD----	KELFSSCMP- 71
Db	22	QDFAADPYTYARLRAGFAHRVT	PEGDEVLLVGVYDRARVLAADPRFSKWN	ETPL 81
QY	72	-QOTSGNSIINMPKHTKIRSVV	NKAFTRVWKQWEPRIQETDEITQKF---	QGR 126
Db	82	TAERALNNHMLSEDEPRHTR	LKLVAREFTWRVLLPRVQEIYDGLVD	AMLA2DGR 141
QY	127	SFEDLVHDSYPLPVIVIS	ELLGVPSAHMEQFKAMSDLLVSTPKDS	ESAFAKFLERDK 186
Db	142	A--DLMESLAWLPFTVISE	LLGVPEPDRAAFRVWTDAFVF--	PDPAQAQTAM----- 191

```
QY 187 CEEELAAFFAGIIEEKNPEQDIISILVE-ABETGEKLSGEBLIPFCTILLVAGNETTT 245
Db 192 --AEMSGYLSRLDSKRGQGDGLLSALVTSDEDSRLTSELLGMAHILLVAGHETT 249
QY 246 NLSINAMYSILETPGVYVEELRSHPELMPQAVEALFRAPA-PVLRRIAKRDTTEIGGHLI 304
Db 250 NLIANGMYALLSHPDQALALRAOMTLLDGAVEEMLYEGVESATYRFPVPEVDLDGTVI 309
QY 305 KEGDMVLAFFASANROEAKFDRPHMEDIRRHNPHTAFGHIFCLGAPLARLEANIALT 364
Db 310 PAGDTVLVWLADAHRTPEFPDRPHEDIRDTAGHLAFGHIFCGIAPLARLEARIAR 369
QY 365 SLISAFPHMECVSITPIE-----NSVIYGLKSRVK 395
Db 370 ALLERCPDL-ALDVSFGLVWYFNPNGIRGLKALPIR 404

RESULT 2
US-09-105-537-39
; Sequence 39, Application US/09105537A
; Patent No. 6265203
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-39

Query Match 31.3%; Score 642; DB 3; Length 416;
Best Local Similarity 37.4%; Pred. No. 2.4e-54;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AVHFFPWYESMRKDAP---VSPDEENQVNSVFLYDDVKVVGDD---KELFSSCMP- 71
Db 22 QDFAADPYPTIYARLRAEGPAHRVTRTEGGEVWLVVGVDRARAVLADPRFSKDRNSTPL 81
QY 72 -QOTSSIGNSIINMDPPKTKIRSVNKAFTPRVMKOWEPRIQEIITDELIOKF-----QGR 126
Db 82 TEAAALNHNMLESDPPRHTLRKLVAEFTMRVVELLRPRVQEIVDGLVDAMLAAPDGR 141
QY 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQKAWSDLLVSTPKDSEAEKAPLEERDK 186
Db 142 A--DLMESLAWPLPIVISELLGVPEPDRAAFRVWTDADFV--PDDPAQAQTAM----- 191
QY 187 CEEELAAFFAGIIEEKNPEQDIISILVE-ABETGEKLSGEBLIPFCTILLVAGNETTT 245
Db 192 --AEMSGYLSRLDSKRGQGDGLLSALVTSDEDSRLTSELLGMAHILLVAGHETT 249
QY 246 NLSINAMYSILETPGVYVEELRSHPELMPQAVEALFRAPA-PVLRRIAKRDTTEIGGHLI 304
Db 250 NLIANGMYALLSHPDQALALRAOMTLLDGAVEEMLYEGVESATYRFPVPEVDLDGTVI 309
QY 305 KEGDMVLAFFASANROEAKFDRPHMEDIRRHNPHTAFGHIFCLGAPLARLEANIALT 364
Db 310 PAGDTVLVWLADAHRTPEFPDRPHEDIRDTAGHLAFGHIFCGIAPLARLEARIAR 369
QY 365 SLISAFPHMECVSITPIE-----NSVIYGLKSRVK 395
Db 370 ALLERCPDL-ALDVSFGLVWYFNPNGIRGLKALPIR 404

RESULT 3
US-09-141-908-13
; Sequence 13, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-141-908-13

Query Match 31.3%; Score 642; DB 4; Length 416;
Best Local Similarity 37.4%; Pred. No. 2.4e-54;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY 21 QD-AVHFFPWYESMRKDAP---VSPDEENQVNSVFLYDDVKVVGDD---KELFSSCMP- 71
Db 22 QDFAADPYPTIYARLRAEGPAHRVTRTEGGEVWLVVGVDRARAVLADPRFSKDRNSTPL 81
QY 72 -QOTSSIGNSIINMDPPKTKIRSVNKAFTPRVMKOWEPRIQEIITDELIOKF-----QGR 126
Db 82 TEAAALNHNMLESDPPRHTLRKLVAEFTMRVVELLRPRVQEIVDGLVDAMLAAPDGR 141
QY 127 SEFDLVHDFSYPPLVIVISELLGVPSAHMEQKAWSDLLVSTPKDSEAEKAPLEERDK 186
Db 142 A--DLMESLAWPLPIVISELLGVPEPDRAAFRVWTDADFV--PDDPAQAQTAM----- 191
QY 187 CEEELAAFFAGIIEEKNPEQDIISILVE-ABETGEKLSGEBLIPFCTILLVAGNETTT 245
Db 192 --AEMSGYLSRLDSKRGQGDGLLSALVTSDEDSRLTSELLGMAHILLVAGHETT 249
QY 246 NLSINAMYSILETPGVYVEELRSHPELMPQAVEALFRAPA-PVLRRIAKRDTTEIGGHLI 304
Db 250 NLIANGMYALLSHPDQALALRAOMTLLDGAVEEMLYEGVESATYRFPVPEVDLDGTVI 309
QY 305 KEGDMVLAFFASANROEAKFDRPHMEDIRRHNPHTAFGHIFCLGAPLARLEANIALT 364
Db 310 PAGDTVLVWLADAHRTPEFPDRPHEDIRDTAGHLAFGHIFCGIAPLARLEARIAR 369
QY 365 SLISAFPHMECVSITPIE-----NSVIYGLKSRVK 395
Db 370 ALLERCPDL-ALDVSFGLVWYFNPNGIRGLKALPIR 404

RESULT 4
US-09-657-440-18
; Sequence 18, Application US/09657440
; Patent No. 6503455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
```

```
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-10

Query Match      29.6%; Score 606.5; DB 4; Length 399;
Best Local Similarity 38.0%; Pred. No. 6.9e-51;
Matches 136; Conservative 66; Mismatches 133; Indels 23; Gaps 7;

QY   29 WYESMRKADPVSFDEENOVSVFLDYDKVKVGVKDLFSSCMPQ---QTSSIGNSLIN-M 84
DB   19 WLREMKHHHPVHEDEYG-AFHVRHADVLVTASDPGVYSQSLSRPPGSQLSEQLSVI 77
QY   85 DEPKHTKIRSVVNKAFTPRVMKWQWEPRIQEITDELIOKEQGRSEFDLVDFSYPLPVI 144
DB   78 DPMHRTLRRLYSQAFTPTVADLEPRVTELAGQLLDVVG-DTFLVADFAYPLPVI 136
QY   145 SHLLGVPSAHMQFKAWSDLLVSTPKDSEAEAKAFLEERDKCEE-----LAA 193
DB   137 AELLGVPPADRILFRSWSDRML-----QMVAADPADMGFGDDADEYQELVKEPMRAMHA 191
QY   194 FPAIGIEEKRNPEQDIISILVEAETGEKLSEELIPFCTLLLVAGNETTNLIISNAMY 253
DB   192 YLHDVTDREARANDLISALVAARVEGERLTDEQIVFEGALLMAGHVSTSMLGNTVL 251
QY   254 SILETPGVYEIRSHPELMPOQAVEALRPAPAVLRRIAKRDTTEIGHLIKEDGMVLAF 313
DB   252 CLKDHPRABAAAARSILPALIEEVLRLRPITTNARVTTKDTVLAGTTIPAGRNVVPS 311
QY   314 VASANDEAKFORFMDFRRHENPHIAFGHGHCGLGAPLARLEANIALTSLSAPP 371
DB   312 LLSANDEOVTFDPHLDJARE-GQIARGHGHICGLGAPLARLEORIALDALFDRFP 368

RESULT 6
US-09-252-991A-17836
Sequence 17836, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: ASBUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17836
LENGTH: 437
TYPE: PR1
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17836

Query Match      28.2%; Score 578; DB 4; Length 437;
Best Local Similarity 33.4%; Pred. No. 4.9e-48;
Matches 131; Conservative 76; Mismatches 147; Indels 38; Gaps 8;

QY   2 NVLNRRQALQRALNGKNKQDAYVHFPEWYESMRKDA---PVSDEBNQVSWVELVDVKK 58
DB   19 NVPDRLKRLGEELIS-----PUHAL--YDGLQVOGAPRAHRAEHPVMMVTVRYRDARK 70
QY   59 VYGD-----KELFSCHMPQQTSSIGNSI-----INMDPKHTKIRSVNKATPTPV 104
DB   71 VLNHGVRDRDAQALYAKRTGPSGAGLEGUSHHMLNDLPDPHTRKSLCVGRAETPRQ 130
QY   105 MKOWEPRIQITDELIOKEQGRSEFDLVDFSYPLPVIISSELLGVPSAHMEQFK-AWS 163
DB   131 VERLQHPIERTIELLDVAGREQADLMADFAIPTAVIFELLGIPEAREHARQSM-- 188
QY   164 LVNSTPKDSEAEAKAFLEERDKCEEBAFFAGIIEBKNEEQODIIILVFAETGEK 223
DB   189 -----ERQALLSPPEAQALADQVDYLAVLLEAKRGQADVVYSGLVQAASEGQ 239

; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-657-440-18

Query Match      31.3%; Score 642; DB 4; Length 416;
Best Local Similarity 37.4%; Pred. No. 2.4e-54;
Matches 148; Conservative 77; Mismatches 137; Indels 34; Gaps 12;

QY   21 QD-AVHPFWYEMRKDAP---VSFDEENOVSVFLDYDKVKVGD---KELFSCHMP- 71
DB   22 QQPADPYTYAKLRAGEGNHRVETTEGEDVWLUVGYDRARAVLADPRSKOWNSTTPL 81
QY   72 -QOTSSIGNSIIMDPPKHTRSVVNKAFTPRVMKWQWEPRIQEITDELIOKF----OGR 126
DB   82 TEAEAALNNMLSDPPRTLKLVAREETMRVVELLRPRVQEIYDGLVDLAALAAPDGR 141
QY   127 SEFIADVDSYPLPVIIVISELLGVPSAHMQFKAWSDLLVSTPKDSEAEAKAFLEEROK 186
DB   142 A--DLMESLAWPUPTIVISELLGVPEPDRAFRVWTDAVF--PDDPAQQTAM----- 191
QY   187 CEELAAFAAGIITEERNPEQDIISILVE-ASETGEKLSEELIPFCTLLLVAGNETTT 245
DB   192 --AEMSGY-LRLIDSXKGQDGEDLLSALVRTSDSGSRLTSELLGMHILLVAGHETT 249
QY   246 NLISNMYSILETPGVYEIRSHPELMPOQAVEALRPAPA-PVLRARIAKRDTTEIGHLI 304
DB   250 NFLANGMYALLSPDQALRALDMTLLDGAVERMLERYEGVESATYRFVPVEDLDGTVI 309
QY   305 KEGDMVLAFVASANDEAKFDRPHMFDIRHPNPHIAFGHGHCGLGAPLARLEANIALT 364
DB   310 PAGTVLVVLADAHRTPERPDPHFRDIRDTACHLAFGHGHCIGAPLARLEANIAR 369
QY   365 SLISAFPHEMCVSITPIE-----NSVIYGLKSFRVK 395
DB   370 ALLERCDDL-ALDVSPGELWYPNPIMRGLKALPIR 404

RESULT 5
US-08-765-907A-10
Sequence 10, Application US/08765907A
Patent No. 6352839
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BMAAS-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSCHE, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marx
APPLICANT: DUTRUC-ROSSET, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
TITLE OF INVENTION: Mutasyntesis
FILE REFERENCE: Streptogramin genes
CURRENT APPLICATION NUMBER: US/08/765,907A
CURRENT FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 10
LENGTH: 399
TYPE: PR1
```

Qy	224	LSGBELIPFCTLLVAGNETTNNLSNAMYSILETPGVYBELRSHPELMPQAVEALRPR	283
Db	240	LSAEVLSMAHLLNMSGFTTWNIGNALVTLLVNPQELALRAQPELLPNAMEELVRHD	299
Qy	284	APAPV-LRRIAKRTQTEIGGHLIKEGDMVLAFVASANROBAEKFDPRHMDIRRHPPHIAF	342
Db	300	SPVTSMLRFTVEVSELDGVTIPAGEYILVSNLTANHDAERDDDRDLDTNTDGHLY	359
Qy	343	GHGHFCIGAPLREANIALTSLISAPHHME	374
Db	360	GFQVHYCWGASLAEGRIALQRIARPDLO	391

RESULT 7

```

US-09-335-409-8
; Sequence 8, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goeblach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-8

```

Query Match 25.3%; Score 518.5; DB 3; Length 419;
Best Local Similarity 32.6%; Pred. NO. 3.1e-42;
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;

Qy	23	AYHPFFWYEMSRKADAPVSFDEENQVWSVFLYDDVKKVVGDKEL-----FSSCM	70
Db	25	AEOPFAIERLEATPIFYMDGEGSWLTRYHDVSAVFRDERFAVSRBEMWESSAYSSAI	84
Qy	71	PQOTSISGNSTINNDPPKHTKIRSVNKAFTPRVMKQWEPRIQETIDELIQKQFGRSFD	130
Db	85	PELSDMKYGLGUPPEDHARVKLVNPFSTRAIDLRLAELQRTVQULDARSQBEED	144
Qy	131	LVHDFSYPPLVIVISELGLVPSAHMEQPKAWSD-----LLVSTPKDKSEAEKAFLEERD	185
Db	145	VVRDYAEGIPMARISALLKVPABCEKXPRRFGSATARGVLVPQVDEETKLVA----	200
Qy	186	KCEELAAFPAGITIEERKNPEQDILSILVBAETGEXKLSEELIPFCFTLLLVAGNETT	245
Db	201	SVTEGLALLHDVLDERRRNPLENDVLTMLLOAEADGSELSTKELVALVCAITAACTDTTI	260
Qy	246	NLISNAMYSILETPGVVVEELRSHPELMPQAVREALRFPAPVLK-----RIAKRTEIGG	301
Db	261	YLIAFANILURSPEALVELVAEPGLMRNALDEVLRP---DNILRIGTVRPARQOLEVCG	317
Qy	302	HLIKEGDMVLAFVASANRDEAKFDPRPHMEDIRRHNPNHIAFGHGHPCLGAPLARLEANI	361
Db	318	ASIKKGEMVLLIFSALRDGTWFSRPOVDFVRRDGTGASLYVGRGHVFCGVSLARLEAEI	377
Qy	362	ALTSLSISAPHEMCVSTTPI	381
Db	378	AVGTIFRRFFPEMK-LKETPV	396

RESULT A

US-09-423-814-71
; Sequence 71, Application US/09413814

```

; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/33535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ. ID. NO 71
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-71

```

Query Match	25.3%;	Score 518.5;	DB 3;	Length 419;
Best Local Similarity	32.6%;	Pred. NO. 3.1e-42;		
Matches 124;	Conservative	72;	Mismatches 155;	Indels 29; Gaps 6

QY	23	AYHPFPWYSMRKXDA	PVSFDEENQWVSFLYDDYDKTKVGD	KEL-----FSSCM	70
		24	AYHPFPWYSMRKXDA	PVSFDEENQWVSFLYDDYDKTKVGD	71
DB	25	AEDFPFAIERLEAT	PIFYDEGRSWLTRYHVS	AVFRDFAVSREWESSAEYSSAI	84
		26	AEDFPFAIERLEAT	PIFYDEGRSWLTRYHVS	85
QY	71	PQOTSSIGNSI	IIMDDPKHFKIRSVNKA	TPTPRMKNQWEPRIQIBELIQFOGRS	130
		72	PQOTSSIGNSI	IIMDDPKHFKIRSVNKA	131
DB	85	PEUSDMMKKYGLF	PEPDEHARVRLVNPST	SRAIDLLRAEIQTVQDILLDARGQEE	144
		86	PEUSDMMKKYGLF	PEPDEHARVRLVNPST	145
QY	131	LVDHDFGVP	LPVIVISELLGVPSAHMEQ	FKAWS-----LIVSTPKDKSEEA	185
		132	LVDHDFGVP	LPVIVISELLGVPSAHMEQ	186
DB	145	VVRDYAEGIP	MRALSALLKYPACDEK	FRFGSATARGVLVPQVDEET	200
		146	VVRDYAEGIP	MRALSALLKYPACDEK	201
QY	186	KCEBELAAP	FAGHIEBKRNKPEODI	SILVBAETGKLSGEHLIPCTLLV	245
		187	KCEBELAAP	FAGHIEBKRNKPEODI	246
DB	201	SVTEGIALLD	VDUDERRNPLENDVIT	MLLQAEADGSRLSKELVALVG	260
		202	SVTEGIALLD	VDUDERRNPLENDVIT	261
QY	246	NLISNAMY	SFLETPGVYEELRSHPE	LMQAVEZALFRAPAPVLR-----	301
		247	NLISNAMY	SFLETPGVYEELRSHPE	302
DB	261	YLIAFVNL	IRSRPEALVELYKAS	PEGLMRNALDEVLRF---DNI	317
		262	YLIAFVNL	IRSRPEALVELYKAS	318
QY	302	HLTKBGDM	VLAFVASANRDBAK	DRPHMFDIRRHENPHIAFGHIG	361
		303	HLTKBGDM	VLAFVASANRDBAK	362
DB	318	ASIKKGBW	FLLTPSALRDGT	VTSRDPVDFVRDCTGASL	377
		319	ASIKKGBW	FLLTPSALRDGT	378
QY	362	ALTSLSA	FAFHEMCVSIT	PI 381	
		363	ALTSLSA	FAFHEMCVSIT	
DB	378	AVGTIER	FERPEMK-LKET	TPV 396	
		379	AVGTIER	FERPEMK-LKET	

BR.SIII.T 9

US-09-568-102-8
 Sequence 8, Application US/09568102
 Patent No. 6346404
 GENERAL INFORMATION:
 APPLICANT: Schupp, Thomas
 APPLICANT: Ligon, James
 APPLICANT: Molnar, Istvan
 APPLICANT: Zirkle, Ross
 APPLICANT: Cyr, Devor
 APPLICANT: Goerlach, Joern
 TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF ECOTHIONES

[illegible]

Db 201 SVTEGLALLHDVDERRRNPLENDVLTMLQAEADGSRSLTKELVALVGAIAGTDTTI 260
Qy 246 NLISNAMYSLTEPGVVEELRSHPELMPQAVEALRFRAPVLR-----RIAKEDTEIGG 301
Db 261 YLIAFAVLNLRSPLEALVKABPGLMRNALDEVLR---DNILRIGTVRPARQDLEYCG 317
Qy 302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHFNPHIAFGHGHIFCLGAPLARLEANI 361
Db 318 ASIKKGEWVLLIPSLALROGTVFSRDPVDFVRDRTGASLAYGRGPHVCPGVSLARLEAEI 377
Qy 362 ALTSLSAPPHEMCVSITPI 381
Db 378 AVGTIFRRFPPEMK-LKETPV 396

RESULT 12

US-09-568-486-8
; Sequence 8, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkie, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-8

Query Match 25.3%; Score 518.5; DB 4; Length 419;
Best Local Similarity 32.8%; Pred. No. 3.1e-42;
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;
Qy 23 AYHPPFWYEMRKDAPVSDVEENQVMSVFLYDDVKKVVGDKEL-----FSSCM 70
Db 25 AEDPPFAIRLEATPIFYWDEGRSWLTYHDVSAVFRDERFAVSREWESESSAEYSSAI 84
Qy 71 PQTSSIGNSIINMDPKTKIRSVNKAFTPRVMQWEPRIQEIITDELQKFGQSEFD 130
Db 85 PELSDMKYGLGLPPEDEHARVKLVNPSFTSRAIDLLRAEIQRTVDQLDARSQBEFD 144
Qy 131 LVHDFSYPPLVIVISSELLGVPSAHMEQFKAWS-----LLVSTPKDSREAEKAFLEERD 195
Db 145 VVRDYAEGIPMRSAISALLKVPACDEKFRFGSATABALGVLPQVDEETKLVA----- 200
Qy 186 KCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGEBELIPFCTLLVAGNETTT 245
Db 201 SVTEGLALLHDVDERRRNPLENDVLTMLQAEADGSRSLTKELVALVGAIAGTDTTI 260
Qy 246 NLISNAMYSLTEPGVVEELRSHPELMPQAVEALRFRAPVLR-----RIAKEDTEIGG 301
Db 261 YLIAFAVLNLRSPLEALVKABPGLMRNALDEVLR---DNILRIGTVRPARQDLEYCG 317
Qy 302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHFNPHIAFGHGHIFCLGAPLARLEANI 361
Db 318 ASIKKGEWVLLIPSLALROGTVFSRDPVDFVRDRTGASLAYGRGPHVCPGVSLARLEAEI 377
Qy 362 ALTSLSAPPHEMCVSITPI 381
Db 378 AVGTIFRRFPPEMK-LKETPV 396

RESULT 13

US-09-568-472-8
; Sequence 8, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkie, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-8

Query Match 25.3%; Score 518.5; DB 4; Length 419;
Best Local Similarity 32.6%; Pred. No. 3.1e-42;
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;
Qy 23 AYHPPFWYEMRKDAPVSDVEENQVMSVFLYDDVKKVVGDKEL-----FSSCM 70
Db 25 AEDPPFAIRLEATPIFYWDEGRSWLTYHDVSAVFRDERFAVSREWESESSAEYSSAI 84
Qy 71 PQTSSIGNSIINMDPKTKIRSVNKAFTPRVMQWEPRIQEIITDELQKFGQSEFD 130
Db 85 PELSDMKYGLGLPPEDEHARVKLVNPSFTSRAIDLLRAEIQRTVDQLDARSQBEFD 144
Qy 131 LVHDFSYPPLVIVISSELLGVPSAHMEQFKAWS-----LLVSTPKDSREAEKAFLEERD 185
Db 145 VVRDYAEGIPMRSAISALLKVPACDEKFRFGSATABALGVLPQVDEETKLVA----- 200
Qy 186 KCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGEBELIPFCTLLVAGNETTT 245
Db 201 SVTEGLALLHDVDERRRNPLENDVLTMLQAEADGSRSLTKELVALVGAIAGTDTTI 260
Qy 246 NLISNAMYSLTEPGVVEELRSHPELMPQAVEALRFRAPVLR-----RIAKEDTEIGG 301
Db 261 YLIAFAVLNLRSPLEALVKABPGLMRNALDEVLR---DNILRIGTVRPARQDLEYCG 317
Qy 302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHFNPHIAFGHGHIFCLGAPLARLEANI 361
Db 318 ASIKKGEWVLLIPSLALROGTVFSRDPVDFVRDRTGASLAYGRGPHVCPGVSLARLEAEI 377
Qy 362 ALTSLSAPPHEMCVSITPI 381
Db 378 AVGTIFRRFPPEMK-LKETPV 396

RESULT 14

US-09-567-899-8
; Sequence 8, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkie, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899

```

; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-8

Query Match      25.3%; Score 518.5; DB 4; Length 419;
Best Local Similarity 32.6%; Pred. No. 3.1e-42;
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;

QY 23 AYHPPVYESMRKDPAPVGFDERNOVNSVFLYDDVKVWGDKEL-----FSSCM 70
DB 25 AEDPPAIERLEATPIFYWDEGRSWLTRYHDVSAVFRDERFAVSRREWESSAYSSAI 84
QY 71 PQOTSSIGNSIINMDPPKHTKIRSVVNKAFTRPVVMKQWEPRIQEITDELIOKPGQRSEPD 130
DB 85 PELSDMKYGLFGLPPEDHARVKLVNPSFTSRALDLRAEIQRTVDQLDARSQGEERFD 144
QY 131 LVHDFSYPLVIVISELGVPSAHMEQKAWSD-----LLNSTPKDKSEAEKAFLEERD 185
DB 145 VVRDYAEGIPRAISALKVPAECDEKFRFGSATARALGVGLVPOVDDEETKLVA---- 200
QY 186 KCEBELAFAFFAGIIEKRNKPEQDIISILVEABETGKLSGELIPFCTLLLVAGNETT 245
DB 201 SVTEGLALLHVDERRRNPENLVITMLLQAEAGSRLSKELVALVGAIIAAGTDTTI 260
QY 246 NLISNAMYSIETPGVYBELRSHPELMPQAVEEALRFRAPAVLR-----RIAKRDEIGG 301
DB 261 YLIAFAVLNLLRSPEALELVKAEPLMRNALDEVLRF---DNILRIGTVRPARQDLEYCG 317
QY 302 HLIKEGDMVLAFAVSAWRDEAKFDRPHMFDIRRHNPNIHAFCHGHIHFCGLGAPLARLE 361
DB 318 ASIKKGMVFLIIPALRDGTVFSRDPVDFVRDRTGASLAYGRGPHVCPGVSLARLEAEI 377
QY 362 ALTSLISAFPHMECVSITPI 381
DB 378 AVGTIFRRFPENK-LKETPV 396

```

```

RESULT 15
5212296-6
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE HARDER, PATRICIA A.;LETO, KENNETH
; J.; O'KEEFE, DANIEL P.;OMER, CHARLES A.;ROMESSER, JAMES A.
; TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:6:
; LENGTH: 406
5212296-6

```

```

Query Match      23.1%; Score 475; DB 6; Length 406;
Best Local Similarity 32.2%; Pred. No. 5.4e-36;
Matches 125; Conservative 68; Mismatches 159; Indels 36; Gaps 10;

QY 19 NKQDAYHPFPWYESMRKDAP-----VSPDEENQVMSVELYDDVKVWGDKEL----- 65
DB 19 NRSCEYQLPDGYAQLR-DTFGPHRVTLVYDGRQAWVTYKHEARKLILGDPRLSSNRDTDN 77

```

```

66 PSSCMP--QOTSSIGNSIINMDPPKHTKIRSVVNKAFTRPVVMKQWEPRIQEIT----DEL 119
DB 78 FPATSPRFEAVRESPOAFIGLDPPEHSTRRMTISFTVKRIKGMREFEVEVHGFDEM 137
QY 120 IQKFQGRSEFDLVHDFSYPLVIVISELGVPSAHMEQKAWSDLLVSTPKDKSEAEKA 179
DB 138 L---AAGPTADLVSOQFALPVPMSVICRLGLGVPIADHEFFQDASKRLV-----OSTDAQSA 189
QY 180 FLEERDKCEELAAFPAGIIEKRNKPEQDIISILVEABETGKLSGELIPFCTLLLLVA 239
DB 190 LTAEND-----LAGYLDGLITQFTEPGAGLVGALVADQLANGEIDREELISTAMLLIA 244
QY 240 GNETTNLISNAMYSIETPGVYBELRSHPELMPQAVEEALRFRAPAVLR-----RIAKRDE 298
DB 245 GHETTASMTSLSVITILLDHPQYAAALRADRSIVPGAVEELLRLVLAADIAGGRVATADIE 304
QY 299 IGHHLIKGDMVLAFAVSAWRDEAKFDRPHMFDIRRHNPNIHAFCHGHIHFCGLGAPLARLE 358
DB 305 VEGHLIRAGEGVIVVNSIANRDGTVYEDPDALDIHRSARHHLAFGFGVHCCLGQNLARLE 364
QY 359 ANIALTSLISAFPHMECVSITPIENSIV 386
DB 365 LEVILNALMDRVPTLRLA--VPVEQLVL 390

```

Search completed: May 28, 2004, 23:25:34
Job time : 93 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 21:28:11 ; Search time 437 Seconds
(without alignments)
12397.578 Million cell updates/sec

Title: US-09-869-334B-2

Perfect score: 1191

Sequence: 1 atgaatgtttaaccgccc.....gcttcctgtgaaatgttaa 1191

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubnpa/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubnpa/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubnpa/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubnpa/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubnpa/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubnpa/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubnpa/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubnpa/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubnpa/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubnpa/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubnpa/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubnpa/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubnpa/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubnpa/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubnpa/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	425	35.7	966	9	US-09-974-300-2190
2	242	20.3	1213	9	US-09-974-300-2182
3	149	12.5	1233	15	US-10-205-032-7
4	149	12.5	60196	15	US-10-205-032-1
5	135.4	11.4	1197	15	US-10-156-761-576
6	135.4	11.4	9025608	15	US-10-156-761-1
7	121	10.2	84428	17	US-10-229-1488-1
8	120.4	10.1	1215	15	US-10-214-446-39
9	117.8	9.9	1197	15	US-10-156-761-2364
10	117.8	9.9	9025608	15	US-10-156-761-1
11	115.4	9.7	1215	15	US-10-156-761-4523
12	114.2	9.6	1164	15	US-10-156-761-6226
13	114.2	9.6	1248	15	US-10-214-446-55
14	112.8	9.5	1290	9	US-09-738-626-617

c	15	112.8	9.5	3309400	9	US-09-738-626-1	Sequence 1, Appli
	16	111.4	9.4	1257	13	US-10-389-647-206	Sequence 236, App
	17	110.2	9.3	1155	9	US-09-974-300-2185	Sequence 2185, Ap
	18	106.2	8.9	1179	15	US-10-156-761-1975	Sequence 1975, Ap
	19	106	8.9	1215	15	US-10-214-446-49	Sequence 49, Appl
c	20	105.4	8.8	82993	15	US-10-080-170-645	Sequence 645, App
	21	103	8.6	215	9	US-09-974-300-2242	Sequence 2242, Ap
	22	97.6	8.2	1191	15	US-10-156-761-7165	Sequence 7165, Ap
	23	96.8	8.1	1371	15	US-10-156-761-3523	Sequence 3523, Ap
	24	95.4	8.0	1278	15	US-10-214-446-19	Sequence 19, Appl
	25	93	7.8	1239	15	US-10-214-446-35	Sequence 35, Appl
	26	92.4	7.8	1209	15	US-10-205-032-9	Sequence 9, Appl
	27	92.4	7.8	1248	15	US-10-156-761-408	Sequence 408, App
	28	92.4	7.8	60196	15	US-10-205-032-1	Sequence 1, Appli
	29	92.4	7.8	100000	15	US-10-156-761-15103	Sequence 15103, A
	30	92.2	7.7	1233	15	US-10-156-761-826	Sequence 826, App
c	31	90.4	7.6	135638	15	US-10-314-657-1	Sequence 1, Appli
	32	89.6	7.5	86941	17	US-10-461-194-2	Sequence 2, Appli
	33	88.8	7.5	1233	10	US-09-953-348-24	Sequence 24, Appl
	34	88.8	7.5	1233	15	US-10-267-255-24	Sequence 24, Appl
	35	88.8	7.5	12249	15	US-09-953-348-74	Sequence 74, Appl
	36	88.8	7.5	12249	15	US-10-267-255-74	Sequence 74, Appl
	37	88.8	7.5	18331	10	US-09-953-348-96	Sequence 96, Appl
	38	88.8	7.5	18331	15	US-10-267-255-96	Sequence 96, Appl
	39	87.6	7.4	71989	13	US-10-282-122A-25888	Sequence 25888, A
c	40	87.6	7.4	71989	13	US-09-939-964-1	Sequence 1, Appli
	41	86.8	7.3	536165	10	US-10-156-761-7447	Sequence 7447, Ap
	42	86.6	7.3	1248	15	US-10-156-761-2153	Sequence 2153, Ap
c	43	85.8	7.2	1263	15	US-10-156-761-2153	Sequence 1, Appli
	44	84.8	7.1	85692	17	US-10-461-194-1	Sequence 22, Appl
	45	84.4	7.1	1224	10	US-09-953-348-22	

ALIGNMENTS

RESULT 1

US-09-974-300-2190
; Sequence 2190, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Herka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/05/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2190
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2190

Query Match 35.7%; Score 425; DB 9; Length 966;
Best Local Similarity 65.4%; Pred. No. 9.7e-27;
Matches 623; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy	230	TTGGAATTCATCATTAACATGACCGCCGCAACATCAAAAATCCGTCAGTCGCA	289
Db	7	TTGGAATTCATCATTAACATGACCGCCGCAACATCAAAAATCCGTCAGTCGCA	66
Qy	290	ACAAAGCCTTTACTCCGCGGTGATGAAGCAATGGAACCGAGATTCAGAAATACACAG	349
Db	67	ACAAAGCCTTTACTCCGCGGTGATGAAGCAATGGAACCGAGATTCAGAAATACACAG	126
Qy	350	ATGAAGTATTCATCAAAAATTTTCAGGGCGCGAGTGTGAGTTTGACCTTGTTCACGATTTTCAT	409

```
Db 127 ATCAATTGCTCGCTGATGTCGCGGAGGAGAGATCGACCTTGTACAGATTTTTCAT 186
Qy 410 ACCGGCTTCGGTTATTTGTAATCTGAGCTGTGGAGTGCCTTCAGCGCAGATGMAAC 469
Db 187 ATCCGTTGCCGTAATCGTCAATTCGGAATTTGTCGGCGTTCCTTTGGTGTATAAGCATC 246
Qy 470 AGTTTAAACATGCTGCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAAGCTG 529
Db 247 ATTTCAGAAATGCTCGACCTGTTGTCAGCTGCGGAGAGCGGATCGCGGAGATG 306
Qy 530 AAAAGACCTTTTGGAAAGACAGATAAGTGTGAGGAAGAACTGGCGCGCTTTTGGCG 589
Db 307 TGAACAGTGGAAAAACATCAGGAGCCAGGAGCCAGGAGAGAGAGCTGACCGCATCTTTGAAA 366
Qy 590 GCATCATAGAAAGAACGGAACAAACCGGAACAGGATATTTCTATTTAGTGAAG 649
Db 367 AGATGATTAAGAGAAAGACAAACCTCGGCAATGACCTGATTTTCGCTTTGATCAAG 426
Qy 650 CGGAAGAAACAGGCGAGAGAGCTGTCGCTGAGAGAGCTGATTCGCTTTTGCACGCTGCTGC 709
Db 427 CCGAACAGAGAGGAGCAAGCTGTCCTGCTGATGAATTTGGTTCCATTTTGCATCTGCTCT 486
Qy 710 TGGTGGCGGAATAGAAACCACTACAAACCTGATTTCAAAATGCGATGTACAGCATATTAG 769
Db 487 TGAATGCGGGGAATGAACACGACGAAATTTAGTTTCAAAATGCGGCTTACAGCATTTCTCG 546
Qy 770 AAACGCGAGCGGTTTACGAGGAACTCGGACGCACTCTGAACCTGATGCTCAGCGAGTGG 829
Db 547 AAACACCGGCGGTATGACGAGCTCGCCAGGCACTCTGAACCTGATTTCCGAGGAGTGG 606
Qy 830 AGGAACCTTTGGCTTTTCAGAGCGCGCGCCCGCTTTTGGAGCGCATTTGCCAACGCGGATA 889
Db 607 AGGAACCGCTTCGCTTTTCGCGGCGCAGCGCGGATGTTGGCGCTTCTTAAACAGGATA 666
Qy 890 CGGAGATCGGGGGACCTGATTAAGAGAGTGAATGTTTGGCGTTTGGCGTGTGGCATCGG 949
Db 667 CGGAGATCAGAGGATTAAGCTTTGAAAGAGAGAGAGTGTGATCGCTTTCTCTTGGCTCTG 726
Qy 950 CAAATCGTATGAAGCAAAAGTTTGAAGAGCGGACATGTTTGTATATCGCGCGCATCCCA 1009
Db 727 CCAACCGTATGAACGAATTTGAGAGGCGCGCAGCAATTTGATATTCACCGCATCCGA 786
Qy 1010 ATCCGATATTTGGTTCGCGCAGCATCCATTTTGGCTTTGGGCGCCCGCTTGGCGGTC 1069
Db 787 ACCCGCATATCGGATTTGGGCAAGCATCCATTTTGGTGGCGCGCGCTCGCAAGGC 846
Qy 1070 TTGAAGCAATATCGGTTAACTGTTTGAATTTCTGCTTTCTCATATGAGTGGGTCA 1129
Db 847 TGAAGCGCGGATCGCATTTGAGCGGCTCTTAAAGCAGTATGCTCGATGGAAGAGCTTG 906
Qy 1130 GTATCACTCCGATTGAAACAGTGTGATATACGATTAAGAGCTTCCGCTGTG 1182
Db 907 CGGTCTGCGGATGCGGACAGCAGCATGTACGCTTGAACATTTCCGCTG 959
```

RESULT 2

```
US-09-974-300-2182
; Sequence 2182: Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10095.500-US
; CURRENT APPLICATION NUMBER: US/09/974, 300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/890,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
```

```
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 2182
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2182

Query Match 20.3%; Score 242; DB 9; Length 1213;
Best Local Similarity 53.6%; Pred. No. 3.7e-67;
Matches 550; Conservative 0; Mismatches 470; Indels 6; Gaps 2;

Qy 75 TCCTTTCCATGCTATGATTCGATGAGAAAGATGCGCCCTGTTCTTCTTGTATGAAGAAA 134
Db 92 TCCTTTCCGATTTATAACCGGATCAGAAAGAGCTCTCTCTGCTGTACCGATGATGAGCG 151
Qy 135 CCAAGTGTGGAGCGTTTCTTTATGATGATGTCAAAAAAGTTTGTTCGGGATAAAGATT 194
Db 152 AAATGCTTTGATATCTTTTCTTATGAGAGCTTCAATTCGTGTTAAAAAACCCGAAAGCT 211
Qy 195 GTTTCCTCAGTTGCTGCGGAGAGCAAGCTCTATTGGAAATTCATCTTAAACATGGA 254
Db 212 CTTCTC---TTCAAAACGCGCGGTANTATGAA--GGAAAAAGTATATTAACAAAGGA 265
Qy 255 CCGCGGAGCATACAAAAATCGTTCAGTCTGTAACAAAGCCCTTACTCGCGCGCTGAT 314
Db 266 CCGCGGAGACACAAAAATGAGAGCCATCGTTAATAAAGCTTTTACCGCGAAGCGGT 325
Qy 315 GAAGCAATGGAAACCGAGAAATCAAGAAATCAAGATGAACCTGATTCAAAAATTCAGGG 374
Db 326 GAAGAGCTTGAACCGCATATCGAGAGTGAAGCTTTTATTAACGAGCGAANCA 385
Qy 375 GCGCAGTGTGATTTGACCTGTTCAAGATTTTCAACCCGCTTCCGCTTATTTGATATC 434
Db 386 GAAAGAAATTTGATGCTGTGGAGCGACTTGGCTGCTCTCTCTCCCTCATTTATCATCGC 445
Qy 435 TGAGCTGCTGGAGAGTGCCTTCAGCGAGATGGAACAGTTTAAAGCATGCTGATCTCT 494
Db 446 TGAATTTTAGCGCTTCCGCTGAAGACCGCTCATGTTTAAACATTTATCAGACATCT 505
Qy 495 GGTCACTACACCAAGATTAAGATGAAGAGCTGAAAAAGCCCTTTTGGAGAACAGGA 554
Db 506 TGTCCGAGGTGCGAGACCGCTCGCTGAGCGCGCGAGCGATGTACAAACGAGCTGA 565
Qy 555 TAAAGTGTAGGAAGAACTGCGCGCTTTTTCGCGCATCATAGAGAAAGCGAAACAA 614
Db 566 AGAAGCAATCGGTTTGTGGCGGATTTTAAACAACTTATTAAGCAGCGCGCAAAAGA 625
Qy 615 ACCGACACAGGATATTTCTATTTTAGTGAAGCGGAAGAACAGCGCAGAGAGCTGTC 674
Db 626 GCCAAAGAGCAGCTGATTTTCGCTTTTACGCGGCGGAAAGTTGACGCGAAATCGCTGAC 685
Qy 675 CGGTGAAGAGCTGATTCGCTTTTTCAGCGCTGCTGCTGGTGGCGCGGAAATGAAACCACTAC 734
Db 686 AGAAGAAAGCTGCTTCACTTTTTCGCTCATCTTTTGGTGCAGGCAATCAGACGCAAC 745
Qy 735 AAACCTGATTTCAAAATGCGAGTACAGCATATTAGAAACCGCGAGGCTTTTACGAGAACT 794
Db 746 CAACTTGATCGCAACACAGCGCTCGCTATCTCACAGAAAGATAAAATCACACGAGAGCGCT 805
Qy 795 GCGCAGCCATCTGAACTGCTCAGCGAGTGGAGAGCCCTTGGTTTCAGAGCGCC 854
Db 806 AAGCAAGATCCGCTCCCTGCTGCTTCTTTTGTGAAGAAATCTGCTGCTTATTTATCGCC 865
Qy 855 GCGCCCGGTTTTCAGCGCATTTGCCAAGCGGATACGAGATCGGGGGCGCCCTGATTA 914
Db 866 CTGTCAAGCGATCGCGCGACGCGGCGCAGAGAGCTTGTATCGGAGCGCTGAGGATTC 925
Qy 915 AGAAGTGTATGCTTTTGGCGTTTGTGGCATCGGCAATCGGCAATCGTGAACAAAGTTGA 974
Db 926 AAAGGGTCTTACAGTGATCAGCTGGGTGCTTCAAGCAATCGTGAACAACTTAAGTTGA 985
Qy 975 CAGACCGCACATGTTGATATCGCGCGCATCCCAATCCCATATTCGCTTTGCGCACGG 1034
```

Db 986 CGATCCTGACAGCTTCAAGCTTGATCGCMATCAAAACCTCATATGAGGCTTCGGCTCGG 1045
Qy 1035 CATCCATTTTGGCTTGGGGCCCGCTTCCCGTCTTGAAGCAAAATATCGCTTAAAGCTC 1094
Db 1046 CATCCATTTTGGCTTGGGGCTTCCCTCGCCGGCTTGAAGCAAAAGTTCGCTCGATTA 1105
Qy 1095 TTTGAT 1100
Db 1106 CTTGCT 1111

RESULT 3
US-10-205-032-7
; Sequence 7, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-20S
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-7

Query Match 12.5%; Score 149; DB 15; Length 1233;
Best Local Similarity 48.5%; Pred. No. 6.5e-37;
Matches 410; Conservative 0; Mismatches 435; Indels 0; Gaps 0;

Qy 239 CCATCAITAAATGACGCGCGGAGCAGATCAAAAAATCGTTTCAGTCTGTAACAAGCCT 298
Db 254 CCATCGCGGCGATCGACCGCGCGGCGAGCGCGCTGCGCAAGCTGTGTGAGCCAGCGT 313
Qy 299 TTACTCCGCGGCTGATGAAGCAATCGGAACCGAGAAATCAAGAAATCAAGATGAAGTGA 358
Db 314 TCACCCCGCGGATCGCCAGATGAGCTGCGATCGGAGATCAACCGCGAGCTGC 373
Qy 359 TTCAAAAATTCAGGGCGCGAGTGTGACCTTTCAGATTTTCAATACCGCTTC 418
Db 374 TCGACCAAGTACGCGGACGAGGACCGGATCGACATCGCCAGCGACCTCGTACCGCTGC 433
Qy 419 CGGTATGTGATATCTGAGTCTGGGAGTGCCTTCAGCGCAGATGAACAGTTTAAG 478
Db 434 CGGTAGCGTATCGCGGAGTCTCGGCAATCCCAAGGATCAAGAGATTCGTAAGCTTCGCG 493
Qy 479 CATGCTCTGATCTTCTGCTAGTACACCGAAGGATAAAAGTGAAGAGCTGAAAAGCCT 538
Db 494 AGTGGTGGACATCATCTCAGCAAGAGGCTGAGTATCCCAACTCCCGGAGACT 553
Qy 539 TTTTGGAGAACGAGATGATGAGGAGAACTGGCGCGCTTTTTCGCGGATCATAG 598
Db 554 TCACCGAGACGCTGGGCGCCCGCTCGAGGAGTGTGTCGGAATTCCTGTAAGCTTCGCG 613
Qy 599 AAGAAAGCGAAACAAACCGGAACAGGATATATTTCTATTATTAGTGAAGCGGAAGAA 658
Db 614 CCCAAGCGCGCGAACCAGGAAGACGATCATAGCGGCTCTGTGCGCGGAGGTCG 673
Qy 659 CAGCGGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTTTGCACGCTGTCTGTGGCGG 718
Db 674 ACGGGCGCAAGCTGACCGAGGAGGTCGTAACATCGTCGCTGTCTCACCAGCG 733
Qy 719 GAATGAAACCACTAGCAACCTGATTTCAATGCGATGTACAGCATATTAAGAACGCCAG 778
Db 734 GGACATATCTCAGCGCGCACGCTGTCTCAGCAACCTGTTCTGCTGTGAGGAGCACCGCG 793
Qy 779 GCGTTTACGAGAACTCGCGAGGCTATCTGACTGATGCTCTCAGCACTGAGGAGCCT 838

Db 794 AGSCACAGCGCGCGTCCCGCGCGAGCCGCGAGCTCTGTCGGCGGCTGATCGAGGAGCGC 853
Qy 839 TGGCTTTAGAGCGCGCGCGCCCGCTTTTCAGGCGCATTCGCAAGCGGATACGAGATCG 898
Db 854 TGGCTACCGCTCCCGTTCACCTGATCTTCGGATCTGACGAGGACACCGACATCC 913
Qy 899 GGGGCACTGATTAAGAAAGTATATGTTTTCGGCTTTCGGCATCGGCAATTCGTT 958
Db 914 TCGGCGCACCCCATGCGCAAGGCGCAGATGCTGATCGCTCGATCGCTCGCGAACC 973
Qy 959 ATGAAGCAAGTTTGACAGCGCGCATGTTGATATCGCGCCATCCCAATCCGATA 1018
Db 974 ACACGAGGTGTTTCAGGACCGCGACACCTTGACATCGACGCGAGTGAACAAGACC 1033
Qy 1019 TTGCGTTTGGCCACGCGCATCTTTTTCCTTGGGCGCGCTTTCGCTTTCGAGCA 1078
Db 1034 TGGCGTTTCGCGACGCGCATCCACCTGCTGGCGCGCTTTCCTGGCAGGCTGAGGCGA 1093
Qy 1079 ATATC 1083
Db 1094 AGGTC 1098

RESULT 4
US-10-205-032-1/c
; Sequence 1, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-20S
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 60196
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-1

Query Match 12.5%; Score 149; DB 15; Length 60196;
Best Local Similarity 48.5%; Pred. No. 1e-35;
Matches 410; Conservative 0; Mismatches 435; Indels 0; Gaps 0;

Qy 239 CCATCAITAAATGACGCGCGGAGCAGATCAAAAAATCGTTTCAGTCTGTAACAAGCCT 298
Db 369 CCATCGCGGCGATCGACCGCGCGGCGAGCGCTGCGCAAGCTGTGAGCGGCT 3549
Qy 299 TTACTCCGCGGCTGATGAAGCAATCGGAACCGAGAAATCAAGAAATCAAGATGAAGTGA 358
Db 3548 TCACCCCGCGGATCGCCAGATGAGCTGCGATCGGCGAGATCAACCGCGAGCTGC 3489
Qy 359 TTCAAAAATTCAGGGCGCGAGTGTGACCTTTCAGATTTTCAATACCGCTTC 418
Db 3488 TCGACCAAGTACCGGACCGGACCGGATCGACATCGCAGCGACCTCGCTACCGCTGC 3429
Qy 419 CGGTATGTGATATCTGAGTCTGGGAGTGCCTTCAGCGCAGATGAACAGTTTAAG 478
Db 3428 CGGTGACGCTCATCGCGGAGTGTTCGCAATCCCAAGGATCAAGAGATTCGCGG 3369
Qy 479 CATGCTCTGATCTTCTGCTAGTACACCGAAGGATAAAAGTGAAGAGCTGAAAAGCCT 538
Db 3368 AGTGGTGGACATCATCTCAGCAAGAGGCTGAGTATCCCAACTCCCGGAGACT 3309
Qy 539 TTTTGGAGAACGAGATGATGAGGAGAACTGGCGCGCTTTTTCGCGGATCATAG 598
Db 3308 TCACCGAGACGCTGGGCGCCCGCTCGAGGAGTGTGTCGGAATTCCTGTAAGCTTCGCG 3249
Qy 599 AAGAAAGCGAAACAAACCGGAACAGGATATATTTCTATTATTAGTGAAGCGGAGAA 658

```
Db 3248 CCCAAGCGCGCGGAAACGAGAGAGCACTGANTCAGCGGCCCTCTGTGCGCGGAGGTGG 3189
Qy 659 CAGCGGAGAGCTGTCCGCTGAAGAGCTGATTCGGTTTTCGACGCTGTCTGTCTGTGGCGG 718
Db 3188 ACGGGCGGAGCTGACCGAGAGGAGTGTCTCAATCGTCTGGCTGTCTGCTCACCGCGCG 3129
Qy 719 GAAATGAACCACTACAAACCTGATTTTMAATGCGGATGTACAGCATATTAGAAACGCCAG 778
Db 3128 GGCACATCTCCAGCGCGACGCTGCTCAGCAACCTGTTCTCTGTGTCTGGAGGAGCACCGCG 3069
Qy 779 GGGTTTACGAGAACTGCGCAGCCATCTCTGAATGATGCTCAGSCAGTGGAGGAGAGCT 838
Db 3068 AGGACAGCGCGCGGTCCCGCGCGAGCGCGAGCTCTGTGCGGCGGATCGAGGAGAGCG 3009
Qy 839 TSGGTTTCAGAGCGCGCGCGCGCGTTTTCAGGCGCATTCGCAAGCGCGGATACGAGATCG 898
Db 3008 TSGGCTACCGTCCCGGTTCACTGATCTCTCGGATCTCTGAACGAGGACACCGCATCC 2949
Qy 899 GGGGGACCTGATTAAGAGGTGATATGGTTTGGCGTTTGGGATCGGCAAAATCGTG 958
Db 2948 TCGGGCACCCCATGCGCAAGGGCCAGATGATGATCGCTGTGATCGCTCCGGAAACCGCG 2889
Qy 959 ATGAAGCAAGTTTGCACAGCGCATCTTTGATATCCGCGCCATCCCAATCCGATC 1018
Db 2888 ACACCGAGGTGTTACGGACCGCGACACCTTTCACATCCGACGCGAGTGAACAAGCAC 2829
Qy 1019 TTGCGTTTGGCCACGCGCATCCATTTTGGCTTGGGGCGCGCGTTCGCCCTCTTGAAGCAA 1078
Db 2828 TGGCGTTCGCGCACGCGCATCCACTGCTTGGCGCGGTTCTTGGCCAGGCTGGAGGCGA 2769
Qy 1079 ATATC 1083
Db 2768 AGGTC 2764

RESULT 5
US-10-156-761-576
; Sequence 576, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 576
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1197)
US-10-156-761-576

Query Match 11.4%; Score 135.4; DB 15; Length 1197;
Best Local Similarity 48.9%; Pred No. 1.7e-32;
Matches 364; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

Qy 344 TCACAGATGAAGTATCAAAATTTTCAGGGCGGAGTGAGTTGACCTTGTTCAGATT 403
Db 347 TGACCAAGGAAGTGGCGGAAGTTTCCAGAGGGGAGGAGGAGATCGACGTCGTGAGGACT 406
```

```
Qy 404 TTTTATACCGCGTTTCCGGTTATTGTGATATCTGAGCTGTCTGGAGTGCCTTTCAGCGCAGA 463
Db 407 TCGCTTACCACTTCCCGGTGACCGGTGATCTGCGGCTGTCTCGGCATCCCGGACAGGACG 466
Qy 464 TGGACAGTTTAAAGCATGTCGTCTTCTTGGTCACTACACCGAGGATATAAGTGAAG 523
Db 467 AGCAGCTCTTCAGGAGTGGACCGATACCTCTGCTGCGTCCGCGGACATCGTCTCCGACCTCG 526
Qy 524 AAGCTGAAAAGCTTTTTCGAAAGACGAGATAAGTGTGAGGAAGAACTGGCGCGCGTTT 583
Db 527 GTGACACCGCCGAAACGGGACCGAGCGCGCGACCGAGCGCAGCAGGAGATGGGCGAGTACC 586
Qy 584 TTGCGCGCATCATAGAAAGAAACGAAACAAACCGGAAACAGGATATATTTTATTTTAG 643
Db 587 TGTGTCAACTTCGCGGAAACAGCGCGCGGTTCGGGCCACCGCGGACATGCTCTCCGACCTCG 646
Qy 644 TGGAGCGGAAAGAAACAGCGGAGAGAGCTCTCGGTGAAGAGCTGATTCGTTTTCGACGC 703
Db 647 TCACGAAACCGGACCGCGCGCGACCGCTCAGCGAGGAGGACCTGGCGGCGGAAACCATCC 706
Qy 704 TGTGCTGTGTCGCGGAAATGAACCACTACAAACCTGATTTTCAAATGCGATGTACAGCA 763
Db 707 TGTGTTTCATCGCGGACACGAGACCAACCGTCAATCTGATCGCAACCGCGGTCTCTCACCC 766
Qy 764 TATTAGAAACCGCAGCGGTTTACGAGAACTTGGCGAGCCATCTGAACTGATGCTCTCAGG 823
Db 767 TGTGTCGCGCGCGGACCAACTGGACCGTCTGCGGAAAGACCGCCAGCTGTCTCGCGGAG 826
Qy 824 CAGTGAGAGAGAGCTTGGCTTTCAGAGCGCGCGCGCGGTTTTCAGGCGCATTCGCAAGC 883
Db 827 CGGTGGAGAACTGTCTGCGCTACGAGCCCGCGTCCACATGCGCGAGCGGTTCCCTCG 886
Qy 884 GGGATACGGAGATCGGGGGGACCTGATTAAGAAAGGTGATATGTTTGGCGTTTGTGG 943
Db 887 TCGACATCGAGCTCGCGGACCAACGATCCCGGTGCGACGTCGCTCATCTGCGCGCTG 946
Qy 944 CATCGCAATCTGTGATGAAGCAACTTTGACAGCGGACATGTTTGTATTCGCGCGCC 1003
Db 947 CCTCGGCGAGCGCGGACCGGATGCGGTTTCAGCGAAACCGGACCGGTTTCGACCCACCGCC 1006
Qy 1004 ATCCCAATCGCATATTTGGCTTTCAGCGGATCCATTTTCCTTGGCGTGGGCGCGCTGG 1063
Db 1007 CGGACAAACGAGCAGCTCGGCTTCGGCAGCGGATATCCACTGTCTCGGCGGCGGACTCG 1066
Qy 1064 CCGCTCTTGAAGCAATATCGGCTT 1088
Db 1067 CCGGTATCGAAGCGGAGCGCGCT 1091

RESULT 6
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
```

```
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      11.4%; Score 135.4; DB 15; Length 9025608;
Best Local Similarity 48.9%; Pred. No. 9.1e-30;
Matches 364; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

QY 344 TCACAGTGAAGTCTTCAAAATTTGAGGGCCAGTGTGACCTTGTTCACGATT 403
DB 737658 TGACCAAGGAAGTGGCGGAAGTGTTCAGGAGGCGAGATCGAGCTGTCACGACT 737717

QY 404 TTTTCATACCCGCTTCCGGTTATCTGTATATCTGAGCTGCTGGAGTGCTTTCAGCGAGA 463
DB 737718 TCGGCTACCCACTGCGCGTGACCGTGTCTGCGCTGCTGGCATCCGACAGGACG 737777

QY 464 TGGACAGTTTAAAGCATGCTGATCTTCTGTGTCAGTACACGAAAGATAAAGTGAAG 523
DB 737778 AGCAGCTTTCAGGACTGGACCGATACCTCTGTCGCTCGGACATCGGCGCCGAAG 737837

QY 524 AAGCTGAAAGACCTTTTTCGAAGAACGAGATATGTGAGGAGACTGCGCGGTTT 583
DB 737838 GTGACACCGCGAAGCGGACCGAGCGCGCGGACCGAGGAGATGGCGCAGTACC 737897

QY 584 TTGCGCGCATCATAGAGAAAGGCGAAACCGGAAACAGGATATTTTCTATTATTAG 643
DB 737898 TGGTCCAACCTGCGGACAGCGCGCGGTGGCGCCACCGGGACATGCTCTCCGACCTCG 737957

QY 644 TGGAAAGGAAAGAAACAGGCGAGAGCTGTCCGCTGAGAGCTGATTCGGTTTTTGGACGC 703
DB 737958 TCAACGAACCGGACCGCGCGCACGGCTCAGCGAGGAGACTGCGCGCGCAACACCATCC 738017

QY 704 TGCTGCTGGTGGCGGGAATGAAACCACTACAACTGATTTCAAAATGCGATGTACAGCA 763
DB 738018 TGCTGTTTCATCGCGGACACGAGACCACTGCTGATCTGATCGCCACGCGTCTCACCC 738077

QY 764 TATTAGAAAGCCGAGCGGTTTACGAGGAACCTGCGAGCAATCTTGAATGATGCTCAGG 823
DB 738078 TGTTCGCGCGCGGACCAACTGGAACCTGTGCGGGAAGACCCCGAGCTGTGCGCGGAG 738137

QY 824 CAGTGGAGGAAGCTTTCGCTTTAGAGCGCGCGCGCGGTTCGAGCGGATTCGCAAGC 883
DB 738138 CGGTGGAGGAAGCTGCTGCGCTACAGCGCGCGCGGTTCACATGCGCGAGCGGTTCCCGCTCG 738197

QY 884 GGGATACGGAGATCGGGGGGCACTCTGATTAAGAAAGTGATGTTTGGCGTTTGTGG 943
DB 738198 TCGACATCGAGCTGCGCGGACCAAGATCCCGGTGSCAGCTCGCTATCTGGCGGTGG 738257

QY 944 CATCGGCAATCTGTATGAAAGCAAAAGTTTACAGACCGGACATGTTTATATCCGCGGCC 1003
DB 738258 CCTCGGCGAGCGCGACCGCGATGCGGTTTACAGAAACCGGATTCGACCCACCCGCC 738317

QY 1004 ATCCCATCGCATATGCGTTTGGCCACGSCATCCATTTTGGGCCCGCGCTTG 1063
DB 738318 CGGACACCGAGACGCTGCGGTTTCGGCAGCGGTATCCACCTGTGCTGCGCGCGCACTCG 738377

QY 1064 CCGGCTCTTGAAGCAAAATATCGCGTT 1088
DB 738378 CCGGTATCGAAGCGGAGCGCGCGCT 738402

RESULT 7
US-10-229-148B-1/c
; Sequence 1, Application US/10229148B
; Publication No. US20040091975A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midecamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/10/229,148B
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((19063)..(20229))
OTHER INFORMATION: ORF26
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((120307)..(21743))
OTHER INFORMATION: ORF25
NAME/KEY: CDS
LOCATION: Complement((21733)..(22527))
OTHER INFORMATION: ORF24
NAME/KEY: CDS
LOCATION: Complement((22534)..(23571))
OTHER INFORMATION: ORF23
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((23555)..(24463))
OTHER INFORMATION: ORF22
NAME/KEY: CDS
LOCATION: Complement((24460)..(25650))
OTHER INFORMATION: ORF21
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((25647)..(26105))
OTHER INFORMATION: ORF20
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((26180)..(27391))
OTHER INFORMATION: ORF19
NAME/KEY: CDS
LOCATION: (2737)..(28983)
OTHER INFORMATION: ORF18
FEATURE:
NAME/KEY: CDS
LOCATION: (2924)..(42779)
OTHER INFORMATION: ORF1
FEATURE:
NAME/KEY: CDS
LOCATION: (42823)..(48657)
OTHER INFORMATION: ORF2
FEATURE:
NAME/KEY: CDS
LOCATION: (48712)..(59802)
OTHER INFORMATION: ORF3
FEATURE:
NAME/KEY: CDS
LOCATION: (59850)..(64556)
OTHER INFORMATION: ORF4
FEATURE:
NAME/KEY: CDS
LOCATION: (64687)..(70365)
OTHER INFORMATION: ORF5
NAME/KEY: CDS
LOCATION: (70365)..(71078)
OTHER INFORMATION: ORF6
FEATURE:
NAME/KEY: CDS
LOCATION: (71113)..(72360)
OTHER INFORMATION: ORF7
FEATURE:
NAME/KEY: CDS
LOCATION: (72400)..(73665)
OTHER INFORMATION: ORF8
FEATURE:
NAME/KEY: CDS
LOCATION: (73694)..(75043)
OTHER INFORMATION: ORF9
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((75899)..(76570))
OTHER INFORMATION: ORF10
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((76602)..(77765))
OTHER INFORMATION: ORF11
NAME/KEY: CDS
LOCATION: (78039)..(79313)
OTHER INFORMATION: ORF12
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((79391)..(81052))
OTHER INFORMATION: ORF13
NAME/KEY: CDS
LOCATION: (81541)..(82356)
OTHER INFORMATION: ORF14
FEATURE:
NAME/KEY: CDS
LOCATION: (82760)..(83362)
OTHER INFORMATION: ORF15
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((83495)..(84142))
OTHER INFORMATION: ORF16
NAME/KEY: CDS
LOCATION: (84329)..(84428)
OTHER INFORMATION: ORF17 (fragment)
US-10-229-1488-1

Query Match      10.2%; Score 121; DB 17; Length 84428;
Best Local Similarity 47.2%; Pred No. 1.6e-26;
Matches 401; Conservative 0; Mismatches 445; Indels 3; Gaps 1;

QY 239 CCATCAITTAACATGGACCGCGGAGCATACAAATAATCGTTACGTCTGTGAACAAGCCT 298
DB 27147 CCATCGGCGCCCTCGACCGCCGCCACGGCCCGATCGCAAGCTGTGCAGCCAGGCGT 27088
QY 299 TTACTCCGCGGTGATGAGCAATCGGACCGAGATTTCAAGAAATCAGATGAATGA 358
DB 27087 TCACCCCGCGCGGATGGCCGTCTGGAAACCCCGCATCGGSCCGTCACTCAGGAGCTCC 27028
QY 359 TTCAAAAATTTTCAGGCGCGCAGTGTGACCTTTGTTTCAGCATTTTTCATACCCGCTTC 418
DB 27027 TCGATCGCGTGGCGGCCAGGAGACCATCGAGTGTGCGGCGACTCTCTCAGCCCTGC 26968
QY 419 CGGTTATTGTGATATCTGAGCTGTGGGAGTGCCTTCAGCGCAGATGGAAAGTTAAAG 478
DB 26967 CGTCATGTCATCGCGAGCTGTGGGCATACCCGTCCGGTGACCGCGATGTGTTCGGCG 26908
QY 479 CATGCTCTGATCTTCTGTGTCTAGTACCCGAGGATAAAGTCAAGNAGCTGAAAAGCCT 538
DB 26907 GGTGGGTGACACCTCTGCTCACCAAGAGGGCGCTGGAGTACCCGAACCTCCCGGCAACT 26848
QY 539 TTTTGGAGAACGAGATAAGTGTGAGGAGAACTGGCCGCGTTTTTTTGGCGGCATCATAG 598
DB 26847 TCAGCGAGACGATCGCCCGCCCTCAAGSAGATGACGACTATCTCTCGCACCAGATCC 26788
QY 599 AAGAAAGCGAAACAAACCGGAAACAGGATATTATTTCTATTTTGTGGAGCGGAGAAA 658
DB 26787 ACGCCAAAGCGGAGGCCCGGTGACGACCTGATCAGCGGCCCTGTGTCCAGGCGCGAGCAG 26728
QY 659 CAGGCGAGAGCTGTCCGCTGAGAGCTGATTCGCTTTGCAAGCTGTCTGCTGGTGGCG 718
DB 26727 ACGGCGGCAAGCTCACCGAGTGGAGATGCTCAACATGCTGGCCCTGTCTGCTCAACGCGCG 26669
QY 719 GAAATGAAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTTAGAAAAGCCAG 778
DB 26667 GCCACGTCTCTCAAGCAGCGCTGCTCAGCAACTCTTCTGTGTACTGGAGGAGACCCCGC 26608
```

```
QY 779 GCGTTTACGAGAACTCGCAGCCCATCTGAATGATGCTCTCAGCAGTGGAGAAAGCCT 838
Db 26607 AGCGCTGGCGGACCTCGCGCGCCGACCGCGAGCTGGTGACCGGTGGAGGAGACAC 26548
QY 839 TGGTTTACAGAGCCCGCCCGCTTTTGGAGCGCATTCGCAAGCGGATACGGAGTCG 898
Db 26547 TGGCTACCGAGCGCCCTTCAACAACATCTTCCGCTTCTCAAGGAGGACACCGACATCC 26488
QY 899 GGGGCGACCTGATTAAGAAGGTGATAGTGTGTTGGGCGTTTGTGSCATCGGCAATCGTG 958
Db 26487 TCGCCCGGGAATGAAGAGGCGCAGATGGTATCGCTCGAGCGCAGTGGCCACCGCG 26428
QY 959 ATGAAGCAAGTTGACAGACCGCACATGTTTGTATATCGCGCCGCGCATCCCA---ATCGCG 1015
Db 26427 ACCCGCAACATCTCCCGAGCGCGACACCTTCGACATCCGACGCTCAAGCAGCTCCCGCC 26368
QY 1016 ATATTGGTTTGGCCACCGGCATCCATTTTCCCTTGGGCGCCCGCTTGCCTGCTTGAAG 1075
Db 26367 ACATGGCGTTCCGGCATCGGTATTCACACTGCTGGCGCCCTTCTTGCACTGAGGAG 26308
QY 1076 CAATATCG 1084
Db 26307 GCAAGGTGG 26239

RESULT 8
US-10-214-446-39
; Sequence 39, Application US/1021446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hightman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 1215
; TYPE: DNA
; ORGANISM: Bacterial
US-10-214-446-39

Query Match 10.1%; Score 120.4; DB 15; Length 1215;
Best Local Similarity 49.6%; Pred. No. 1.3e-27;
Matches 418; Conservative 0; Mismatches 391; Indels 33; Gaps 3;

QY 241 ATCATTAACATGACCGCGCGGAGCATACAAAATCCGTTGCTGCTGAAACAAAGCCTTT 300
Db 277 ATCTCTAGCATGACCGCGCGGAGCATACACCCCGGTGGCGACCCGCTGGTGGCCAAAGCGTTC 336
QY 301 ACTCCGCGGTGTAAGCATGGACCGGAGATTCAAGAAATCAGAGATCAACTGATT 360
Db 337 ACCATGACCGAGGTGGAGAGTTGCGCCCGCGGTGGCGAGCTGGCGAGAGCTGATC 396
QY 361 CAAAAATTTACGGGGCGGAGTGA---GTTTGACCTTGTTCAGATTTTTCATACCGGCTT 417
Db 397 GACAAGATGTTCCGACCGCGCCCGCTGCACTGGTTCGAGAGTTTCGCGCTGCGGTG 456
QY 418 CCGGTTATGTGATATCTGAGCTGCTGGAGAGTGCCCTTCAGCGCAGATCGAAACAGTTTAAA 477
Db 457 CCGGTGCGGGTGTATCTGCACTGCTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGG 516
QY 478 GCATGCTGATCTTCTGCTCAGTACACCGGAGGATTAAGAGAGCTGAAAAAGCC 537
```

```
Db 557 GCGTGGAGCGAGCGCGCTGTCCACAGTTCCCTGACGCGCGGAGGAGTTGACGCCAC 576
QY 538 TTTTGGAGAAACAGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTTTCCCGGCATCATATA 597
Db 577 C-----AGGAGAACTGCGGCGCTACATCGGSGGTTGATC 612
QY 598 GAAGAAAGCGAAACAAACCGGACAGGATATTTCTATTTTAGTGAAGCGGAAGAA 657
Db 613 GAGGATCACCGGCGCGTCCGCTGAGGACCTGATCACCGGCTGATCGAGCCCGGAGAC 672
QY 658 ACAGCGAGAGAGCTGTCGCGTGAAGAGCTGATTCGTTTTCACGCTGCTGCTGCTGCTG 717
Db 673 CGCGACCGCGCTGACCGGACGAGTGGTGGACCTGCTGCTGCGGATCTCTGTTGGCC 732
QY 718 GGAATGAAACCACTACAAACCTGATTTTCAAAATGCGATGATACAGCATATTTAGAAACGCCA 777
Db 733 GGCACGAGACCAACCGCCACGAGATCCCACTTCGTTGGTGAAGCTGCTGSAACCGGCC 792
QY 778 GCGGTTTACGAGGAACTCGCGGAGCGCATCTGAACTGATGCTCAGGAGTGGAGGAGCC 837
Db 793 GAGCAGTGAACCGGCTGCGGAGGAGACCGGAGCTGGTCCCGACCGCGGTTCGAGGAGCTG 852
QY 838 TTGCGTTTACAGAGCGCGCGCCCGCTT-----TTGAGCGCATTCGCAAGCGGATAGC 891
Db 853 ATGCGTTTTCGCGCTGGCGAGCGCTGCTGCTCCCGGTAAGCCACCGAGGACGTG 912
QY 892 GAGATCGGGGGCGACCTGATTAAGAAAGTGTGATATGTTTGGCGTTTTTGGCATCAGCA 951
Db 913 GAGTTCGCGCGCACGCTGTCGCGCGCGGAGCGCGTCTGTTGGCTCGGCGCGGCC 972
QY 952 AATGCTGATGAAGCAAGTTTGAAGAGCGGACATGTTTGTATATCGCGCGCATCCGAT 1012
Db 973 AACCAGCGACCGCGCGAGTTTCGACGCGCGGAGGAGCTGGACCTCGCGCGGAGGCGCAC 1032
QY 1012 CCGCATATTGCTTGGCGACCGCATTCATTTTCCCTTGGGCGCCCGCTTCCCGCTCTT 1071
Db 1033 CAGACCTCGGTTTCGCGCATGGCGTCCACCACTGCTCGGCGCGCGCTGCGCGCTG 1092
QY 1072 GA 1073
Db 1093 GA 1094

RESULT 9
US-10-156-761-2364
; Sequence 2364, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2364
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1197)
US-10-156-761-2364
```


Query Match 9.9%; Score 117.8; DB 15; Length 1197;
Best Local Similarity 47.9%; Pred. No. 8.8e-27;
Matches 406; Conservative 0; Mismatches 432; Indels 9; Gaps 2;

245 TTAACATGACCCGCGAAGCATACAAATAATCGTTTACGTGTGAGCAAAAGCCTTTACTC 304
239 TGACACGAGCGCGCCGCAACACGCGCTCGTGTGTCTCGACCGCTTCACCG 298
305 CGCGGTGATGAGCAATGGGACCGAGAAATCAAGAAATACAGATGAATTCATAA 364
299 CCCGCGGTGAGGCGCTCGCGCACAGTGCAGCATACAGACGAGCTCTCTCGACA 358
365 AATTTCAGGCGCGCAGTGTGAGTTTGCCTTTTACAGATTTTTCATACCGCTTCGGTTA 424
359 CGATCGTCCCGCGCGAGCGAGCTGTATCGGCGACTTCGCTTCCGCTGGCGATCA 418
425 TTGTGATATCTGAGCTGTGAGAGTGCCTTACGCGAGTGAAGATGAAGATGATGT 484
419 CGGTGATCTGAACTGTGAGAGTGCCTTACGCGAGTGCCTTACGCGAGTGTTC 478
485 CTGATCTTCTGTCAGTACACGAGAGTAAAGTGAAGATGAAGATGAAGATGATGT 544
479 CGAA-----GGAATTCCGGGCGTGAAGACCGACTCCCGCGAGCGCGGTG 532
545 AAGAACGAGATGATGAGGAGAACTGGCGCGTTTATTTTGTAGTGAAGGAGAAAG 604
533 ACGCGCGCGGTGAGTGTGCGGAGTCTCTGAAATATCTGACCGGACTGTGTGCA 592
605 AGCGAAACAAACCGGACAGGATATATTTTATTTTGTAGTGAAGGAGAAAG 664
593 GCGGTGAGACCGCGCGGAGTGTGCGAGTGTGCGGCGCGCGGAGAGAGAG 652
665 AGAAGCTGTGCGGTGAAGAGTGTGCGTGTGTCAGCGTGTGCTGTGCGGAGAA 724
653 ACCGCTGAAAGAGCGCGAGTGTGCTGTCCATGATGTCCTGTGCTGTGCGGCT 712
725 AAACCATACAAACCTGATTTTCAATGAGATGATGAGCAATATTAAGAACCG 784
713 AGACGCGTCAATCTGATCGGCAACGCGACCTGCGCTGTGCGGCGACCCCG 772
785 ACGAGGAATCGGAGCGCATCTGTAAGTGTGCTGAGTGTGAGGAGAGAGCTT 844
773 TCGGCTGTGCGGAGCGACCGGAACTGTGAGTGTGCGGCTGTGAGGAGATG 832
845 TCAGAGCGCGCGCGGTGTGAGGCGCATTTGCAAGCGGAGTACG---GAGAT 901
833 ACGACGGTGTGAGAGCGGAGCGTGTGCGGTTCGCGCTCGAACCCATCGAG 892
902 GGCATCTGATTAAGAGGTGATGTTGTTGCGGTGTGCGCATCGGCAATTCGT 961
893 GCAAGCGCATGAGAGGCGCAACCGGTGTCTCTGTGCGCATCGGCAACCG 952
962 AAGCAAGTTTGACAGACCGGACATGTTGATATCGCGCGCATCCCAATCG 1021
953 GGGGAGTTTCGGGACCGGACGACTTCGAGTCAACCGCGCGGACCCCGCG 1012
1022 CGTTTGGCAGCGCATCCATTTTTCCTTGGGCGCGCGCTTCCCGCTCTTGA 1081
1013 CTTTGGGCGGCGTGGCACTTCTGCTCGGTGCTTCCCTGCGCGCGGTGAG 1072
1082 TCGGTT 1088
1073 TCGGTT 1079

RESULT 10

US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 9.9%; Score 117.8; DB 15; Length 9025608;
Best Local Similarity 47.9%; Pred. No. 4.8e-24;
Matches 406; Conservative 0; Mismatches 432; Indels 9; Gaps 2;

245 TTAACATGACCCGCGAAGCATACAAATAATCGTTTACGTGTGAGCAAAAGCCTTTACTC 304
2902578 TGACACGAGCGCGCCCGNACACACGCGCTCGTGTCTCGACCGCTTCACCG 2902519
305 CGCGGTGATGAGCAATGGGAAACCGAGAAATCAAGAAATACAGATGAATTCATAA 364
2902518 CCAGCGCGGTGAGGCGCTGCGCGCACAGGTGCAGCATACAGCGAGCGCTCTCGACA 2902459
365 AATTTCAGGCGCGCAGTGTGAGTTTGCCTTTTACAGATTTTTCATACCGCTTCG 424
2902458 CGATGTCGCCCGCGGCGAGCTGTATCGGCGACTTCGCTTCCCGCTGGCGATCA 2902399
425 TTGTGATATCTGAGTGTGCGGAGTGCCTTTCAGCGCAGATGAAACAGTTTAAAGCA 484
2902398 CGGTGATCTGTAATGCTGTGAGTGTGCGGAGTGCCTTCCAGCGGACAGAGCTGT 2902339
485 CTGATCTTCTGTCTAGTACACCGAAGGATTAAGAGTGAAGAGCTGAAAGCCTTT 544
2902338 CGAA-----GGAATTCCGGGCGTGAAGACCGACTCCCGCGGCGGCTG 2902285
545 AAGAACGAGTAACTGTGAGGAGTGTGCGCGCTTTTTCGCGCATCATAGAGAA 604
2902284 ACGCGCGCGCGTGTGAGTGTGCGCGCTTCTCGAATATCTGACCGGCTGTGCGACA 2902225
605 ACGGAAACAAACCGGAAACAGGATATTTTCTATTTAGTGAAGCGGAGAAACAG 664
2902224 GCGTTCAGACCCGCGCGAGCGACTGTGCGCGCTGTATCGCGCGCGGAGAGCG 2902165
665 AGAAGTGTGCGGTGAGAGCTGATTCGTTTTCGCTGTGCTGTGTGCTGTGCTG 724
2902164 ACCGCTGAACAGGCGGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 2902105
725 AAACCACTACAACCTGATTTCAATGCGATGATACAGATATAGAAACCGCGGCTT 784
2902104 AGACGAGTCAATCTGATCGGCAACCGGACCTTGGCGCTTGTGCGCGCACCAAC 2902045
785 ACGAGGAATCGGCGAGCGCATCTGAACTGATGCTTCAAGAGTGAAGAGCTT 844
2902044 TCGCGCTGTGCGCGAGCGACCGGAACTGTGAGACTCGCGCTGTGAGGAGATG 2901985
845 TCAGAGCGCGCGCGCGCTTTTGAAGCGCATTTGCGAGCGGAGATAG---GAGAT 901
2901984 ACGACGGTGTGTCGAGACCGGAGCGTGTGCGGTTCGCGCTCGAAACCATCGAG 2901925
902 GGCATCTGATTAAGAGGTGATGTTTGGCGTTTGTGCGCATCGGCAATCTGTATG 961


```
Db 291 CACCCTGGCTCGCTCBAACAACTCCGGAGCCCGTACGGAGACCGCCACCGGCTTCT 350
Qy 360 TCARAAATTTACGGGCGCAGTGAGTTTACCTTGTTCACGATTTTTCATACCCGCTTCC 419
Db 351 CGACCGACTCGGGAGCGACCGGAGACCGGCTCATCGCTCGTACCGCGCGCCCTGCG 410
Qy 420 GGTATTGTGATCTAGCTCTGGAGTGTCTTTCAGCGCAGATGGAACAGTTTAAAGC 479
Db 411 GATCACCGTTCATCTGACCTCTCGGCTGCGGACGAAACCGCGCGGACTTCCGGCG 470
Qy 480 ATGCTCTGATCTTCTGCTCAGTACACCGGAGGATTAAGTGAAGAGCTGAAGAGCTT 539
Db 471 CTGGACCGACCGCTCGTCAACCCCGACCGCGCCCGCGGAGCTCGCAGGGAATCCGT 530
Qy 540 TTTCGAAGAAACAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTGTCCGGCATATAGA 599
Db 531 CGTCTCGT-----TGCTCGCTTCTTTCACCGGCTCTCGC 566
Qy 600 AGAAAGGGAACAAACCGGAACAGATATATTTCTATTTTGTGGAAGCGGAAGAAC 659
Db 567 CGACAAAGCGCAAGAACCCCGCGGACGACCTGCTCTCGGACCTCATCGCGTGCAGGA 626
Qy 660 AGCGCAGAGCTCTCGGCTGAAGAGTGAATTCGTTTTCAGCTGCTCTGCTGTCGCGG 719
Db 627 GGGGACCGGCTCACCGAGGACGAGTGAATGCTTCCCTGCTTCTCTCTCTGCGCGG 686
Qy 720 AAATGAACCACTACAAACCTGATTTCAATTCGATGTACAGCATATTAGAAACGCGAG 779
Db 687 CTRGAGAAACCGTGACCTTCTCGGAAACCGCTACTGGGCTGCTGCGGCATCCGA 746
Qy 780 CGTTTACGAGGAACGCGGAGCCTCTGAACTGAGTCTAGGAGTGAAGAGAGCTT 839
Db 747 CGAGCTCGCGCGCTCCGCGAGGACCGGACCGCTGCGGAGCGCGTCCGCGGAGTTCG 806
Qy 840 CGCTTTCAGAGCGCGCGCCCGGTTT---TGAGGCGCATTCGACGCGGATACGAGAT 896
Db 807 CGCTTACGAGGCGCGCGCTCTCGCATCCCGCTTCCCGGTGCGGAGCTGACGAT 866
Qy 897 CGGGGCGCACCTGATTAAAGAGGTGATATGTTTGGCGTTTGTGGCATCGGCAATCG 956
Db 867 CGCGGGGTCACTCGCGCGCGGGGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 926
Qy 957 TGATGAACAAAGTTTGAACAGCGGACATGTTTGAATTCGCGCGCATTCGCAATCCGA 1016
Db 927 TGACCGGAGCGGTTTCCCGGACCCGACCGGCTCGATCTCGGCGCGGAGCGCGGCG 986
Qy 1017 TATTGCGTTGCGCACGCGATCCATTTTTCCTTGGCGCGCGCTGCGCGCTTGAAGC 1076
Db 987 CTTGCGCTCGGCGCGGCGGCTTCACTTCTGCGCTGGCGCGCGCTGCGCGCTGAGAC 1046
Qy 1077 AAATATCGC 1085
Db 1047 GGAGTGGC 1055
```

RESULT 13

US-10-214-446-55

; Sequence 55, Application US/10214446

; Publication No. US20030180742A1

; GENERAL INFORMATION:

; APPLICANT: Weiner, David

; APPLICANT: Burk, Mark J.

; APPLICANT: Hitchman, Tim

; APPLICANT: Pujol, Catherine

; APPLICANT: Richardson, Toby

; APPLICANT: Short, Jay M.

; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING

; FILE REFERENCE: 09010-500001

; CURRENT APPLICATION NUMBER: US/10/214,446

; CURRENT FILING DATE: 2002-08-05

; PRIOR APPLICATION NUMBER: US 60/309,497

; PRIOR FILING DATE: 2001-08-03

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 55

; LENGTH: 1248

; TYPE: DNA

; ORGANISM: Bacterial

; US-10-214-446-55

Query Match 9.6%; Score 114.2; DB 15; Length 1248;

Best Local Similarity 46.7%; Pred. No. 1.3e-25;

Matches 399; Conservative 0; Mismatches 453; Indels 3; Gaps 1;

Qy 230 TTGAAATTCATATTAACATGAGCCCGCGGAGCATCAAAATATCCGTTAGTGTGA 289

Db 263 TCAGAACTGGCTCGTCTTCTTGGACCCCGCGGACACGCGGCTGCGCGCCAGGTG 322

Qy 290 ACAAGACCTTTACTCGCGCTGATGAAGCAATGGAAACCGGAATTCAGAAATCACAG 349

Db 323 CCCCOCGCTGAGCGCTCCGCGCTACGGGCGCTTGGCTCCCGCGTACGGGAGATCGCG 382

Qy 350 ATGAACCTGATTCAAAAATTCAGGGGCGCAGTTCACCTTGTTCACGATTTTTCAT 409

Db 383 AGAACTCGTACGGCCCTTCGCGCGCGCGCGCTGGTGAACCTGGTCGAGGCTTCGCG 442

Qy 410 ACCCGTTCGCGTTATTTGTAATCTGAGTCTGCGGAGTGCCTTCAGCGCAGATGGAAC 469

Db 443 CCCCCTTCGCGTGTGCTGCTGCGGCGGCTGCTCGGGTTCGACCCCGCTGCGTGGCG 502

Qy 470 AGTTTAAAGCATGCTGATCTTCTGTCAGTACACCGAAGGATAAAGTGAAGAGCTG 529

Db 503 GGTTCGGGAGAGGCG---CTCGCCCTCCAGCGCTCCGGCGGACCGAGGCGACCGGT 559

Qy 530 AAAAGACCTTTTGGAAAGAACGAGATAAGTGAAGAGAACTGGCGCGCTTTTTCGCG 589

Db 560 CGCGGCGCGCTCGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 619

Qy 590 GATCATGAAGAAAGCGAAACAAACCGGAACAGGATATTTCTATTTAGTGGAG 649

Db 620 CGAGCTGCG 679

Qy 650 CGGAAGAAACCGCGGAGAGCTGTCGGTGAAGAGCTGATTCGCTTTTCACGCTCTGC 709

Db 680 CGCGCGCGGAGACCGCTCGCTGGGAGACCGCGCTGACGCTGACCTGCGCTCCACCTCC 739

Qy 710 TGGTGGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATAG 769

Db 740 TGACGCGCGGACGAGACGACGAGCGGCTCTGCGGCAAGCGGCTGCTCGCGCTCG 799

Qy 770 AAACGCGAGCGTTTACGAGGAACTGCGGAGCGGCTGCGGAGCGGCTGCGGAGCG 829

Db 800 CGCGCGCGGAGTGGCGGAGAACTCGCGCGGAGACCGCGGCGCTTTCGGAACGCGCG 859

Qy 830 AGAAAGCTTTCGCTTTCAGAGCGCGCGCGCGCTTTTGAAGCGCATTCGCAAGCGGATA 889

Db 860 ACAGTTCCTGCGCGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 919

Qy 890 CGGAGATCGGGGCGACCTGATTTAAAGAGGTGATATGTTTGGCGTTTGTGGCATCGG 949

Db 920 CGGAGCTCGCGCGCGGAGCGGCTCGCGCGGCGGAGCGCGGTCAGCTGCTGGTTCGG 979

Qy 950 GAAATCGGTGATGAAGCAAGTTTGAACAGCGGACATGTTTGAATTCGCGCGCGCATCCCA 1009

Db 980 CCGACCGGACCGCGCGCGCTTCCGCGGACCGCGCGCGGCTGAGCATTCGCGCGGACCG 1039

Qy 1010 ATCCGATATTCGCTTGGCGACGCGCATCTTTTTCCTTGGCGCGCGCGCTTGGCGGTC 1069

Db 1040 GCGCGCATCGCGCTTCGCTTGGCGATCCACTACTGCTGGCGCGCGCGCTGGCGCGCG 1099

Qy 1070 TTGAAGCAAAATATCG 1084

Db 1100 CGGAGCGGAGATCG 1114

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 20:10:15 ; Search time 82 Seconds
(without alignments)
8060.323 Million cell updates/sec

Title: US-09-869-334B-2
Perfect score: 1191
Sequence: 1 atgaatgtttaaccgccc.....gcttcctgtgaaatgtaa 1191

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	139.2	11.7	8478	3	US-08-676-818-1
2	139.2	11.7	8478	4	US-09-407-549-1
C 3	118.2	9.9	4403765	3	US-09-103-840A-2
C 4	118.2	9.9	4411529	3	US-09-103-840A-1
5	116.4	9.8	4403765	3	US-09-103-840A-2
6	116.4	9.8	4411529	3	US-09-103-840A-1
7	111.4	9.4	1194	4	US-08-765-907A-9
8	111.4	9.4	4496	4	US-08-765-907A-6
9	109.8	9.2	1314	4	US-09-252-991A-1265
C 10	109.8	9.2	1482	4	US-09-252-991A-1338
11	109.8	9.2	1671	4	US-09-252-991A-1222
12	103.6	8.7	6085	3	US-09-029-603-4
13	88.8	7.5	1233	4	US-09-266-965-24
14	88.8	7.5	12249	4	US-09-266-965-74
15	88.8	7.5	18331	4	US-09-266-965-96
16	87.6	7.4	71989	4	US-09-443-501A-2
C 17	86.8	7.3	536165	4	US-09-214-808-1
18	85.8	7.2	1221	6	5212296-16
19	85.8	7.2	1879	6	5212296-5
20	84.4	7.1	1224	4	US-09-266-965-22
21	84.4	7.1	1227	3	US-09-385-028-23
22	84.4	7.1	1227	4	US-09-726-614-23
23	84.4	7.1	1227	4	US-09-385-040-23
24	84.4	7.1	15079	3	US-09-385-028-1
25	84.4	7.1	15079	4	US-09-726-614-1
26	84.4	7.1	15120	4	US-09-385-040-1
27	84.2	7.1	68750	3	US-09-335-409-1

28	84.2	7.1	68750	4	US-09-568-102-1
29	84.2	7.1	68750	4	US-09-567-969-1
30	84.2	7.1	68750	4	US-09-568-480-1
31	84.2	7.1	68750	4	US-09-568-486-1
32	84.2	7.1	68750	4	US-09-568-472-1
33	84.2	7.1	68750	4	US-09-567-899-1
34	81.2	6.8	1407	4	US-09-252-991A-15866
C 35	81.2	6.8	1731	4	US-09-252-991A-15928
C 36	81.2	6.8	2502	4	US-09-252-991A-15767
37	80.4	6.8	47981	4	US-09-679-279-1
38	72.2	6.1	1251	3	US-09-105-537-38
39	72.2	6.1	5970	3	US-09-320-878-21
40	72.2	6.1	5970	4	US-09-141-908-11
41	72.2	6.1	5970	4	US-09-657-440-21
42	71.2	6.0	50937	3	US-09-428-517-1
43	68	5.7	1735	1	US-08-102-863-10
44	68	5.7	1735	5	PCT-US92-10885-10
45	65	5.5	1242	3	US-09-068-132-1

ALIGNMENTS

RESULT 1
US-08-676-818-1
; Sequence 1, Application US/08676818
; Patent No. 6057136
; GENERAL INFORMATION:
; APPLICANT: Bewer, Stanley Grant
; APPLICANT: Perkins, John B.
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Pero, Janice G.
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
; TITLE OF INVENTION: SUSTILIS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,818
; FILING DATE: 08-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/239,430
; FILING DATE: May 6, 1994
; APPLICATION NUMBER: 08/084,709
; FILING DATE: June 25, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 04599/CC4001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8478
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-676-818-1

Query Match 11.7%; Score 139.2; DB 3; Length 8478;

TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
 TITLE OF INVENTION: SUTILIS
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 50Z or 55SX
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: WordPerfect (Version 5.1)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/407,549
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/239,430
 FILING DATE: May 6, 1994
 APPLICATION NUMBER: 08/084,709
 FILING DATE: June 25, 1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Freeman, John W.
 REGISTRATION NUMBER: 29,066
 REFERENCE/DOCKET NUMBER: 04599/034001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8478
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-09-407-549-1

	Query Match	11.7%	Score 139.2;	DB 4;	Length 8478;
	Best Local Similarity	50.4%;	Pred. No.	4.6e-35;	
	Matches	426;	Mismatches	0;	Gaps 3;
QY	252	GGACCCGCGGAAGCATCAAAAATCCGTTCAGTCTGTGAACAAGGCCTTTACTCCGCCGCT	311		
Db	5735	GAACCAGCCTGATCATATAGACGATTGCGAAGCTTGCCACGGAGCGCTTTACGCCGAGAAC	5794		
QY	312	GATGANGCAATGGAAACCGAGAAATCAAGAAATCACAGATGAACATACTGATTCAAAAATTCCA	371		
Db	5795	GACAGAGAGTTATCAGCCGTATATCAATTAAGAACCTGTCCATCATTTGCTTTGATCAAAGTGCA	5854		
QY	372	GGGGCGCAGTGAAGTTTGACCTTGTTCCACGATTTTCATACCCCGCTTCGCGTTATTGTGAT	431		
Db	5855	AGGTAAAAAAMAGATGGAGGTCAATTTCCGACITTCGTTTTCTTAGCAAGTTTGTGAT	5914		
QY	432	AUTCAGTCTGTGGGAGTGCCTTTACGGCAGATGGAAACAGTTTAAAGCATGTCGTGATCT	491		
Db	5915	AGCTAACATTTATAGTGTATACCGGAGGAAGATAGGAGGACAAATTAAGGAGCTGGCTCGAG	5974		
QY	492	TCTGGTCAGTACACCGAAGATAAAGTCAAGAGCTGAAAAGCCTTTTTCGAAGAAG	551		
Db	5975	TTCTCAT-----TCAARCGATGTGATTTTACC CGCTCAAGAAAGGCATTAACAGAGGGCA	6027		
QY	552	AGATAAGTGTGAGGAAGAACTGGCCGCGTCTTTTTCGCCGCATCATAGAAGAAAAGCGAAA	611		
Db	6028	ATATTATGCTGTGCAGGCTATG--GCATATTTCAAGAGCTGATTCAAAGAGAAAACG	6385		
QY	612	CAAACCGGAACGATATATTCTATTTTATGTGAAGCGGAAGAAACAGCGGAGAACT	671		
Db	6086	CCACCCTCAACAGGATATGATCAGCATGCTCTTTGAAGGGGAGAGAAA---AGGATAAGCT	6142		
QY	572	GTCCGCTGAGAGCTCATTCCTTTTGGCAAGCTGTCTGCTCCGCCCCGAAAAAAAAAAGCAG	731		

Db 6143 GACGAAGAGGAGGCGGCATCTACGTGSCATATTGCTGGCGATCGCGGACATGAGACAC 6202
Qy 732 TACAAACCTGATTTCAATGCGCATGTACAGCATATTAGAAACGCCAGGCGTTTACGAGA 791
Db 6203 GGTCAATCTCATCAGCAATTCAGTCCCTTGTCTGTGAGCATCCAGAACAGCTTTTGA 6262
Qy 792 ACTGCGAGCCCATCTGAACTGATGCTCAGGAGTGGAGGAGGCTTCCGTTTCAGAGC 851
Db 6263 ACTGAGAGAAATCCAGATCTTATTGTTACCGCAGTCCGAGGAATGTTTACGCTATGAA 6322
Qy 852 GCGGCGCGCGTTTGGCGCATTCGCCAGCGGATCGAGATCGGGGCGACCTGAT 911
Db 6323 CCCACCGCAATGACGCCAGAGTGGGTGAGAGGATATTGACATCTCGGGGGTGAAGAT 6382
Qy 912 TAAAGAGTGATATGTTTGGCGTTTGTGCGATCGGCAATCTGTGATGAAGCAAGTT 971
Db 6383 CCGTCAAGGAGAACAGTCTATCTTTGTTAGAGCGGTATTCAGACCTTAGCATATT 6442
Qy 972 TGACAGACCGCATGTTTGTATATCGGCGCCATCCCAATCCGATATTGCGTTTGGCA 1031
Db 6443 CACGAACCCGATGTCCTTCGATATTACGAGAAAGTCTTAATCCGATCTTCAATCGGCA 6502
Qy 1032 CGGCATCCATTTTGCCTTGGGCGCCGCTTCCCGCTCTTGAAGCAATATCGCGTTAC 1091
Db 6503 TGGCATCATGTTTGTCTAGGTCTCTGCTGGCACGATTAGAGCGCAAAATGCGATTAA 6562
Qy 1092 GTCTTT 1097
Db 6563 CACTCT 6568

RESULT 3

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007-00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 9.9%; Score 118.2; DB 3; Length 4403765;
Best Local Similarity 48.1%; Pred. No. 2.2e-26;
Matches 401; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

Qy 241 ATCAATTAATGACCGCGGAGCATACAAAATCCGTTGAGTGAACAAAGCCCTTT 300
Db 3948025 ATGATGACATGACGATCCGACATCTGTGGGCGAAGCTGTTTAAAGCGCGCTTC 3947966
Qy 301 ACTCCCGCGTGTGAGCAATGGACCGGATTCAGAAATCAGAGTGAATGAT 360
Db 3947965 ACCGCAAGCGGGTGAAGGACAAAGGCGGTCGATTCGCGGCTGTGTGACACCTGATC 3947906
Qy 361 CAAAAATTCAGGCGCGCAGTGAGTTGACCTTCTCAGCATTTTTCATACCCGCTCCG 420
Db 3947905 GACGCGGTGTGCGACGCGGAGTGAGCTTCTGTGGGACCTGTGGCGCGCCGCTACCG 3947846

Qy 421 GTTATTGTGATATCTGAGCTGTGGAGTGCCCTTCAGCGCAGATGGAACAGTTTAAAGCA 480
Db 3947845 ATGGCGGTGATCGGCGACATGCTCGGGGTGCGTCCAGAGCAGCGGACATGTTCTTGGCG 3947786
Qy 481 TGGTCTGATCTTGTGTCAGTACACCGAAGGATAAAAAGTGAAGAGTGAAGAAAGCCCTTT 540
Db 3947785 TGGTCCGACGATCTGTGTA-CATTCTCAGTTTGCATGTGTCTCAAGAGGATTTCCAGAT 3947727
Qy 541 TTGGAGAACAGAGATAAGTGTGAGGAGAACTGGCGCGGTTTTTTCGCGGCATCATAGAA 600
Db 3947726 CACCATGAGCGCTTCGCGGCGCTTACACGACATTCACCGG-----GCCACCATGG 3947675
Qy 601 GAAAGCGAAACAAACCCGAAACAGGATATTATTCTATTATTAGTGAAGCGGAAGAAACA 660
Db 3947674 GCAAGCGAGCGGACCCACCGACGCTGTGTCAGCGTGTGTGATTCGAAAGTTGAC 3947615
Qy 661 GCGGAGAACTGTGCGGTGAAGAGCTGATTCGCTTTTTCACGCTGTCTGTGTGGCGGA 720
Db 3947614 GCGGAGCGGCTAAGCGACGACGAGCTGTGTCATGGAAGCTGTCTGATCTGATCGGCGG 3947555
Qy 721 AATGAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTAGAAAGCCAGGC 780
Db 3947554 GACGAGACCAACCGCGCATACCTTGAGCGGTGTACCGAGCAGCTGTGCGCAACCGTGAC 3947495
Qy 781 GTTTACGAGAACTGCGCAGCCATCTGAACTGATGCTCAGCAGTGGAGGAAGCCTTG 840
Db 3947494 CAGTGGGACCTGTGACGCGGACCGCTGTGCTGCGCGGCGCATGAGGAGTGTCTA 3947435
Qy 841 CTTTTCAGAGCGCGCGCGGCTTTTGAAGCGCATTTGCAAGCGGATACGAGAGTGGG 900
Db 3947434 CGTGGACCGCGCGGTAAAGAACATGTGCGGGTGTGACCGCGATACCGAGTTCCAC 3947375
Qy 901 GCGCACCTGATTAAAGAGGTGATATGTTTGGCGTTTGTGGCATCGGCAATCGGAT 960
Db 3947374 GGCACGCGTGTGTGCGCGCGAGAGATGATGTCTCTCGAGTCGCGGAACTTCGAC 3947315
Qy 961 GAAGCAAGTTTGACAGACCGCACATGTTTGATATCCGCGCCCATCCCAATCCGATATT 1020
Db 3947314 GAGGCGGTTTCTGTGAACCGGAAAGTTTGAATTCAGGAAATCCHAAACAGCCACTTG 3947255
Qy 1021 GCGTTTGCCACGCGCATCATTTTTCCTTGGCGCGCGCTTGGCGCTTGA 1073
Db 3947254 GCGTTTGGCTTGGCAGCGCATTTCTGCTGGGCAATCAGTGGCGCGGTGGA 3947202

RESULT 4

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007-00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 9.9%; Score 118.2; DB 3; Length 4411529;
Best Local Similarity 48.1%; Pred. No. 2.2e-26;
Matches 401; Conservative 0; Mismatches 423; Indels 9; Gaps 2;


```
QY 729 CACTACAAACCTGATTTCAATGCGATGTACAGATATTAGTAAACCCGAGGCGTTTACGA 788
DB 872199 CACCCTGTAATCTCAATGCGCGGCGTCTGCGCTGGCCGAGAACCTTGACCACTGCA 872258
QY 789 GGAACTGGCGAGCCATCTGAACTGATGCTCAGGAGTGGAGGAGACCTTGGTTTCAG 848
DB 872259 AACGCTGGGAACGATTTTGAAGTTGTGCGCATGCGATCGAAGAGATCGTGAAGTGGAC 872318
QY 849 AGCGCGCGCGCGGTTTGTAGCGCATTGCCAAAGCGGATACGAGATCGGGGGGCACT 908
DB 872319 GTCCGCTGATCACCATCGAAGCGCGGACGCGCTCCGCGGTGAGCTGGCGGCGCGC 872378
QY 909 GATTAAGAGGTGATATGTTTGTGCGTTTGTGCGATCGGCAAAATCGTGATGAAGCAAA 968
DB 872379 GATCGAGCGGCTCAGAAAGTTGTGCTGTGGAGGCGCTCGGCCAACCGTGATCCCGAGCT 872438
QY 969 GTTTGACAGACCGCATCATGTTTGTATCGCGCGCATCCCAATCGCATATTTGGTTTGG 1028
DB 872439 GTTCAGCGCGGCTCAGAAAGTTGTGCTGTGGAGGCGCTCGGCCAACCGTGATCCCGAGCT 872498
QY 1029 CCAGCGCATTCATTTTGTGCGTTGCGGCGCGCGCTTGCCTGCTTGA 1074
DB 872499 TCAGGCGGTGACATATTGCTCGGCGCGCATCTGGCTCGGCTGGA 872544

RESULT 7
US-08-765-907A-9
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAWAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-9

Query Match 9.4%; Score 111.4; DB 4; Length 1194;
Best Local Similarity 48.5%; Pred. No. 2.2e-26;
Matches 415; Conservative 0; Mismatches 416; Indels 24; Gaps 3;

QY 249 CATGACCCCGCGAGCATACAAAATCCGTTCACTGCTGGAACAAAGCTTTACTCCGCG 308
DB 228 CATGACCCCGCGATGACCCGACCCCTGCGCGCTGTGACCGAGGCTTACACCCCGCG 287
QY 309 CGTGATGAAGCAATGGGACCGAGAAATTCAGAAATCAAGATGAATGATTCAGAAATTT 368
DB 288 CACCGTCCCGACCTCGAACCAGCGGTCCACGAACTGGCGCGGCACTGCTCGACGCGGT 347
QY 369 TCAGGGCGGAGTGATTTGACCTGTTTCAGATTTTTCATACCGCTTCCGCTTATGTT 428
DB 348 CGACCGGAGCA---CGTTGACACCTGTGTGCGGACTTCGCTACCGCTGCGCGTATCGT 404
QY 429 GATATCTGAGCTGTGGAGTGCTTTCAGTCAGTCAGATGGAACAGTTTAAGAGTATGTTCTGA 488
DB 405 GATCGCGAATCTCTCGCGGTGCGCGCGCGCGACCGACCCCTGTTCTCGCTCTGTTCCGA 464
```

```
QY 489 TCTTCTGGTCAGTACACCG-----AAGGATAAAAGTGAAGAGTGA 530
DB 465 CCGGATGCTGCAGATCAGGTGCGCGACCGCGGACATCAGTTGCGCGACGACGCGCA 524
QY 531 AAAGGCTTTTGGAAAGACGAGATAGTGTGAGGAGAACTGCGCGCGTGTGTTTTCGCGG 590
DB 525 CGAGGACTACCAACGCTGCTCAAGAAACCCATGCGCATGACGCGCTTACCTCCACGA 584
QY 591 CATCATAGAAGAAACGGAACAAACCGGACAGGATATTTCTATTTTACTGGAAGC 650
DB 585 CACAGTCAACGACCGCGCGCGCGCGCGCGCGCGGAAAGACCTGATCTCGGCACTCGTGGCGC 644
QY 651 GGAAGAAACAGCGGAGAGCTGTCCGTTGAAGAGCTGATTTCCGTTTTTGCACGCTGTGCT 710
DB 645 CCGGTTGAGGAGGAGACGACTCACCGACGAGAGATCGTGGATTCGGGGCGCTGCTGCT 704
QY 711 GGTGGCGGAAATGAAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTAGA 770
DB 705 GATGCGCGGCGACGCTCTCCACCTCCATGCTGTGCGCAACACCGTCTGTGCTGAAGA 764
QY 771 AACGCCAGCGGTTTACGAGGAACTGCGGAGCCATCTGAACTGATGCTCAGGACAGTGA 830
DB 765 CACACCGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGATCGA 824
QY 831 GGAAGCTTTGCTTTTCAGAGCGCGCGCGCGCGGTTTTGAGCGCGATTCGCCAAGCGGATAC 890
DB 825 AGAAGTACTGCGCTGCGGCGCGCGCGATCACCGTTCATGGCGCGGTCCACCAAGGACAC 884
QY 891 GGAGATCGGGGGGACCTGATTAAGAAAGTGATATGTTTGGCGTTTGTGTCATCGCG 950
DB 885 CGTCTCGCGCGCACCCACCATCCCGCGGAGCGCATGCTGTCGCTGCTGCTGCTGCTG 944
QY 951 AAATCGTGATGAAGCAAAAGTTTGACAGACCGGACATGTTTGATATCCGCGCGCATCCCA 1010
DB 945 CAACACGAGCAACAGGTTTTCACGACCGCGGACCGACCTCGACCTCGCGCGGAAAGC-- 1002
QY 1011 TCCGATATTTGCTTTGGCCACGCGCATTCATTTTTCGCTTGGGGCGCGCGCTTGCCTGCT 1070
DB 1003 -CGCCAGATCGCTTTGCGGCGCGCGCATCCACTACTGCTGCGGCGCGCGCTGCGCGCT 1061
QY 1071 TGAAGCAATATCGC 1085
DB 1062 GGAGGCGCGCATCGC 1076

RESULT 8
US-08-765-907A-6
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAWAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Mutasynthesis
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-6
```


Db 1068 TCTCGGCTACGGCTTGGCGTGCACACTACTGCTGGCGCCTCGCTGCGCCCGCTGGAGGG 1127
Qy 1077 AATATCGCGTTACGCTTTGATTTCCTCTTCATATGGAGT 1123
Db 1128 GCGATCGCCATCAGCGCTGCTCGCGGCTTCCCGACCTCCAGT 1174

RESULT 10
US-09-252-991A-1338/c
; Sequence 1338, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1338
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1338

Query Match 9.2%; Score 109.8; DB 4; Length 1482;
Best Local Similarity 47.6%; Pred. No. 8.4e-26;
Matches 422; Conservative 0; Mismatches 432; Indels 33; Gaps 2;

Qy 240 CATCATATACATGACCGCGGAGCATACAAATCCGTTCACTGCTGAACAAAGCCTT 299
Db 1189 CATGCTCAACCTGACCGCGGAGCATACAAATCCGTTCACTGCTGAACAAAGCCTT 1130

Qy 300 TACTCCGCGGCTGATGAAGCAATGGGACCGAGAAATCAAGAAATCAAGATGAATGAT 359
Db 1129 CACCCCGCGCAGTGGAGCGCTGCAACCGCATATAGAACGATACCGAGGAATGCT 1070

Qy 360 TCAAAATTTTACGGGCGCGAGTTGAGCTTGTTCACGATTTTTCATACCGCTTCC 419
Db 1069 GGACGCGCATGCGCGCGCGGAGCAACGCGCGAGCTCATGCGCGAGTCCGCTGAC 1010

Qy 420 GGTATTGTGATATCTCAGTCTGCTGGAGTGCCTTTCAGCGCAGATGGAACAGTTTAAAGC 479
Db 1009 CATCGCGGTGATCTTCAGTCTGCTGGAGTGCCTTTCAGCGCAGATGGAACAGTTTAAAGC 950

Qy 480 ATGCTCTGATCTTCTGCTCAGTACACCGAAGGATATAAAGTGAAGAGCTGAAAAAGCCTT 539
Db 949 GTCTTGGAGCGCCAG-----GCGGAATGCTGTC 920

Qy 540 TTTCGGAAGACGAGATAGTGTGGAAGAACTGGCGCGGTTTTCGCGGCATCATA 599
Db 919 GCGGAGGAGGCGCGCGCGCTGGCGATGCGCAGTCCGAGTACCTACCTGCTGCTCGA 860

Qy 600 AGAAAGCGAAACAAACCGGAAACAGGATATATTTCTATTAGTGAAGCGGAAGAAC 659
Db 859 GGCACAGCGCGCGAGCCACCGACGACGCTACAGCGGCTGTGTGACGCGCGCGCAGA 800

Qy 660 AGCGGAGAGCTGCTCCGCTGAAGAGCTGATTCGCTTTTTCACGCTGCTGCTGGTGGCGG 719
Db 799 GAGCGGCAACTGAGCGAGCGGGAACCTGCTCTCCATGCGCCACCTGTGATGAGCGG 740

Qy 720 AATGAAACCATACAAACCTGATTTCAATGCGATGTACAGCATATATAGAAACGCGCAGG 779
Db 739 CTTTCGAGACCAACCATGATGATGCGCAACGCGCTGCTCACTGCTGCTGCTCAACCGGGA 680

Qy 780 CTTTACGAGGACTGCGGACCGCATCTGAGTGCCTCAGCAGTGGAGGAGCCTT 839
Db 679 GCAACTGGCTTGTGCGGCGCAGCGGAACTCTCTGCGCAACGCGCATGGAAGACTGGT 620

Qy 840 GCGTTTCAGAGCGCGCGCCCGCTTTTTCAG-----GCGCATTCGCAAGCGGGATACGGAGAT 896
Db 619 CCGCCACGACAGCGCGCTGCGCGCTCGATGTTGCGCTTCACCGTGAAGAGCTGGAACT 560

Qy 897 CGGGGGGCACTGATTAAAGAGGTGATATGTTTGGCGTTTGGCGCATCGGCAATCG 956
Db 559 GGACGGGTACCATTTCCCGCGCGGAATACATCTGCTCTCCAACTGACGCGCCACCA 500

Qy 957 TGATGAAGCAAGTTTGACAGACCGCACATGTTTGATATCCGCGCCCATCCCAATCCCA 1015
Db 499 CGACGCGAGCGCTTCGACGATCCGACCGCTCGACCTACCGCAACACCGATGGCCA 440

Qy 1017 TATTGCGTTTGGCAGCGCATCCATTTTTCGCTGGGCGCGCTTGCCTGCTTGAASC 1076
Db 439 TCTCGGCTAOCGCTTCGCGCTGACACTACTGCTGCGCGCTCTGCTGGCGCGCTGGAGGG 380

Qy 1077 AATATCGGCTTACGCTTTTGATTTCGCTTTTCTCATATGAGT 1123
Db 379 GCGGATCGCCATCCAGCGCCTGCTCGCGGCTTCCCGACCTCCAGT 333

RESULT 11
US-09-252-991A-1222
; Sequence 1222, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1222
; LENGTH: 1671
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1222

Query Match 9.2%; Score 109.8; DB 4; Length 1671;
Best Local Similarity 47.6%; Pred. No. 9.1e-26;
Matches 422; Conservative 0; Mismatches 432; Indels 33; Gaps 2;

Qy 240 CATCATTAACATGACCGCGGAGCATACAAATCCGTTCACTGCTGTAACAAAGCCTT 299
Db 253 CATGCTCAACCTGACCGCGCGGACCATACCGCTGCGCTGCGTGGCGCGGCTT 312

Qy 300 TACTCCGCGCTGATGAAGCAATGGGACCGAGAAATCAAGAAATCAAGATGAATGAT 359
Db 313 CACCCCGCGCAGTGGAGCGCTGCAACCGCATATAGAACGATACCGAGGAATGCT 372

Qy 360 TCAAAATTTTACGGGCGCAGTGTGACCTTGTTCACGATTTTTCATACCGCTTCC 419
Db 373 GGACGCGATGCGCGCGCGCAACGAGCGCACTGATGCGGACCTTCGCGATCCGCTGAC 432

Qy 420 GGTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAGC 479
Db 433 CATCGCGTGTATCTTCGAGCTGCTGGCATTTCCGAGCGCGAGCGCAACGCGCGCCA 492

Qy 480 ATGCTCTGATCTTCTGCTCAGTACACCGAAGGATAAAGTGAAGAGCTGAAAAAGCCTT 539
Db 493 GTCTTGGAGCGCCAG-----GCGGAATGCTGTC 522

Qy 540 TTTCGGAAGACGAGATAGTGTGAGGAAGAACTGCGCGCGTTTTCGCGGCATCATAGA 599
Db 523 GCGGAGAGGCGCGCGCGCTGCGCGATGCGCAGTTCGACTACCTGCGCTGCTCGA 582

Qy 600 AGAAAGCGAAACAAACCGGAAACAGGATATTTTCTATTATTAGTGAAGCGGAAGAAC 659

583 GGCCAAAGCGCGGAGCGAGCGCTCTACAGCGGGCTGTGTCAGCGCGCGCGA 642
660 AGCGAGAGCTGTCGGTGAAGCTGATTCCTGTTTGGACGCTGCTGCTGGTGGCGG 719
643 GAGCGGCAACTGAGCGAGGCGGAATCGTCTCCATGGCCACCTGCTGATGATGAGCGG 702
720 AAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAGAAACGCCAGG 779
703 CTTGAGACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 762
780 CGTTTACAGAGCTGCGAGCGATCTGAACTGATGATGATGATGATGATGATGATGATGAT 839
763 GCAATGCGCTGTTGCGGGGCGAGCGGAACTCGTCTCCATGGCCACCTGCTGATGATGAG 822
840 GCGTTTACAGCGCGCGCGCGCTTTTTCAG---GCGATTCGAGCGGATGATGATGATGAT 896
823 CCGCACGACGCGCGGCTGCGGCTCGATGTTGCGCTTTCACCGTGGAGAGCTGGAAT 882
897 CGGCGGCGACCTGATTAAGAGGTGATGATGTTTGGCTTTTGGCATCGGCAAAATCG 956
883 GGAAGGCTACCAATTCGCGCGGGAATACATCTGGTCTCAACCTGACCGCCAA 942
957 TGATGAAGCAAGTTTGAAGAGCGGCAATGTTGATTCGCGCGCATCCCAATCCGCA 1016
943 CGAGCGGAGCGCTTCGAGCATCCGAGCGGCTCGACCTCACCGCAACACCGATGGCCA 1002
1017 TATTGCGTTTGGCGACGCGATCCATTTTTCGCTTGGGCGCGGCTTTCGCTTTGAAGC 1076
1003 TCTGCGTACGCTTTCGCGGCTGCTACTGCGTGGCGGCTGCTGCGCGGCTGAGG 1062
1077 AAATATCGGTTTAACTGCTTTGATTTTCGCTTTTTCCTCATATGAGT 1123
1063 GCGGATGCCATCCAGCGCTGCTCGCGGCTTCCCGACCTCCAGT 1109

RESULT 12
US-09-029-603-4
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Toupet, Christine
; APPLICANT: Pospiech, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; EARLIER FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 6085
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (378)..(1665)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (1747)..(2553)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (2593)..(4011)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (4013)..(4999)
; OTHER INFORMATION: ORF

FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (5071)..(6085)
; OTHER INFORMATION: ORF
US-09-029-603-4
Query Match 8.7%; Score 103.6; DB 3; Length 6085;
Best Local Similarity 51.0%; Pred. NO. 2.3e-23;
Matches 244; Conservative 0; Mismatches 234; Indels 0; Gaps 0;
QY 623 AGGATATTATTCTATTTTACGTGAAGCGGAGAACAGGCGGAGAGCTGTCCGCTGAAG 682
Db 1048 AGGATCTGCTACCTCTCTCTGTCGCGCGCGGAGCACCGGATCACCGCTCAGCGTGGAG 1107
QY 683 AGCTGATTCGCTTTTTCACGCTGCTGCTGGTGGCGGAAATGAAACCACTACAAACCTGA 742
Db 1108 GCATGCTGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1167
QY 743 TTTCAATGCGATGTACAGCATATTAGAAACGCCAGCGGCTTTTACGAGGAACTGCGGAGCC 802
Db 1168 TCGCCAGGCGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1227
QY 803 ATCTGGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 862
Db 1228 CACCGAGTGTGACACCGCGCGCGTGTGAGAGCTGATGCGGTACGACCGCGCGCTGCGAGG 1287
QY 863 TTTTGAAGCGCATTCGCAAGCGGATACGAGATCGGGGGGCGCTGATTAAAGAGGTG 922
Db 1288 CGTACGCGCTGCGGCTGCGGATGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1347
QY 923 ATATGCTTTTGGCTTTTGTGGCATCGGCAATTCGATGAGGAACTGAGCAAAAGTTTACAGACCGC 982
Db 1348 GCGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1407
QY 983 ACATGTTGATATCCGCGCATCCCAATCCGATATTCGCTTTGGCCACGCGATCCATT 1042
Db 1408 AGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1467
QY 1043 TTTGCTTTGGGCGCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 1100
Db 1468 ACTGCTGCGCGGACCTTGGCT 1525

RESULT 13
US-09-266-965-24
; Sequence 24, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-24
Query Match 7.5%; Score 88.8; DB 4; Length 1233;
Best Local Similarity 51.1%; Pred. NO. 6.8e-19;

```
Matches 238; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 641 TAGTGAAGCGGAGAAACAGCGGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTGCA 700
DB 665 TCCTGACCGCGCGGACGAGGAGGACAGCTCAACGAGGGGAGCTGGTGGCTGGCGG 724
QY 701 CGCTGCTGCTGGTGGCGGAAATGAACCACTACAAACCTGATTTCAAAATCGGATGACA 760
DB 725 AGGCCATCTCTCATCGCCGCTACGAGACCTCGGCGAGCCAGATCCCCAACTTCCTCTACG 784
QY 761 GCATATTAGAAACCGCAGGCGTTTACGAGGAACCTCGCAGCCATCCTGAACTGATGCTTC 820
DB 785 TCCTCTTCGCGCCACCCGACGCTGTCTGGAGCGGATCAGGAACGACACGACCTCATCCCCG 844
QY 821 AGGCAGTGGAGGAGCCTTGGCTTTTCAGAGCGCGCGGCCCGCGTT-----TTGAGGGCGCA 874
DB 845 ACGCGCTCGAGGAACCTGTGGCTTCGTGCGCCATCGGCACCGCTGACCGGCTTTCGCCGTA 904
QY 875 TTGCAAGCGGATPACGAGATCGGGGGGACCTGATTAAGAGAGGTGATATGTTTGG 934
DB 905 CGGCCACCGAGGAGCTCGAGCTCGGGGGGAGTCTCTGGTCAGGGCGGGGAGAGCGGTCTGTC 964
QY 935 GCTTTGGCATCGCAATCGTATGAACAAAGTTTGACAGACCGGCACATGTTTGATA 994
DB 965 CGTCGATGGGCGCGCCCAACCGACCCCGAGCTGTTTACGAGACCCCGACGAGCTGGACC 1024
QY 995 TCCGCGCCCATCCCAATCCGCAATATGCTGTGGCCACGCGCATCAATTTTSCCTTGGGG 1054
DB 1025 TCGCGCGCGCGCGAATCGCACTCGGCTTCGGCGCGGGACCGCACCTGCTGGGCG 1084
QY 1055 CCGCGCTGCGCGCTTTGAACCAATATCGGCTTAACTCTTTGAT 1100
DB 1085 CCAACTGGCGCGGCTGGAGCTCCAGATCAGCTCACGACGCTGTT 1130

RESULT 14
US-09-266-965-74
; Sequence 74, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: He, M
; APPLICANT: Varoglu, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT FILING DATE: 1999-03-12
; EARLIER FILING DATE: 1996-08-19
; EARLIER FILING DATE: 1994-10-06
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 12249
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-74

Query Match 7.5%; Score 88.8; DB 4; Length 12249;
Best Local Similarity 51.1%; Pred. No. 2.9e-18;
Matches 238; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 641 TAGTGAAGCGGAGAAACAGCGGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTGCA 700
DB 9239 TCGTGACCGCGCGGACGAGGAGCAAGCTCACCGAGCGGAGCTGGTGGCTGGCGG 9298
QY 701 CGCTGCTGCTGGTGGCGGAAATGAACCACTACAAACCTGATTTCAAAATCGGATGACA 760
DB 9299 AGGCCATCTCTCATCGCCGCTACGAGACCTCGGCGAGCGAGATCCCCAACTTCCTCTACG 9358
QY 761 GCATATTAGAAACCGCAGGCGTTTACGAGGAACCTCGCAGCCATCCTGAACTGATGCTTC 820
DB 9359 TCCTGACCGCGCGGACGAGGAGCAAGCTCACCGAGCGGAGCTGGTGGCTGGCGG 9418
QY 701 CGCTGCTGCTGGTGGCGGAAATGAACCACTACAAACCTGATTTCAAAATCGGATGACA 760
```

```
DB 9299 AGGCCATCTCTCATCGCCGCTACGAGACCTCGGCGAGCGAGATCCCCAACTTCCTCTACG 9358
QY 761 GCATATTAGAAACCGCAGGCGTTTACGAGGAACCTGCGAGGCAATCCTGAACTGATGCTTC 820
DB 9359 TCCTCTTCGCGCCACCCGACGCTGTCTGGAGCGGATCAGGAACGACACGACCTCATCCCCG 9418
QY 821 AGGCAGTGGAGGAGCCTTGGCTTTACAGAGCGCGCGGCCCGCGTT-----TTGAGGGCGCA 874
DB 9419 ACGCGCTCGAGGAACCTGTCTGGCTTCGTGCGCCATCGGCACCGTGGAGCGCTTTCGCCGTA 9478
QY 875 TTGCAAGCGGAGATCGAGATCGGGGGGACCTGATTAAGAGAGGTGATATGTTTGG 934
DB 9479 CGGCCACCGAGGACGTCGAGCTCGGGGGAGTCTCTGGTCAGGGCGGGGAGAGCGGTCTGTC 9538
QY 935 GCTTTGGCATCGCAATCGTATGAACAAAGTTTGACAGACCGGCACATGTTTGATA 994
DB 9539 CGTCGATGGGCGCGCCCAACCGACCCCGAGCTGTTTACGAGACCCCGACGAGCTGGACC 9598
QY 995 TCCGCGCGCATCCCAATCCGCAATATGCTGTGGCCACGCGCATCAATTTTSCCTTGGGG 1054
DB 9599 TCGCGCGCGCGCGCAATCCGACCCCTGGGCTTGGCGCGGGACCGCACCTGCTGGGCG 9658
QY 1055 CCGCGCTGCGCGCTTTCGAACCAATATCGGCTTAACTCTTTGAT 1100
DB 9659 CCAACTGGCGCGGCTGGAGCTCCAGATCAGCTCACGACGCTGTT 9704

RESULT 15
US-09-266-965-96
; Sequence 96, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT FILING DATE: 1999-03-12
; EARLIER FILING DATE: 1996-08-19
; EARLIER FILING DATE: 1994-10-06
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 96
; LENGTH: 18331
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-96

Query Match 7.5%; Score 88.8; DB 4; Length 18331;
Best Local Similarity 51.1%; Pred. No. 3.8e-18;
Matches 238; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 641 TAGTGAAGCGGAGAAACAGCGGAGAGCTGTCCGGTGAAGAGCTGATTCGGTTTGCA 700
DB 9239 TCGTGACCGCGCGGACGAGGAGCAAGCTCACCGAGCGGAGCTGGTGGCTGGCGG 9298
QY 701 CGCTGCTGCTGGTGGCGGAAATGAACCACTACAAACCTGATTTCAAAATCGGATGACA 760
DB 9299 AGGCCATCTCTCATCGCCGCTACGAGACCTCGGCGAGCGAGATCCCCAACTTCCTCTACG 9358
QY 761 GCATATTAGAAACCGCAGGCGTTTACGAGGAACCTCGCAGCCATCCTGAACTGATGCTTC 820
DB 9359 TCCTCTTCGCGCCACCCGACGCTGTCTGGAGCGGATCAGGAACGACACGACCTCATCCCCG 9418
QY 821 AGGCAGTGGAGGAGCCTTGGCTTTACAGAGCGCGCGGCCCGCGTT-----TTGAGGGCGCA 874
```

Db 9419 ACGCCGTCGAGAACTGCTGCGCTTCGTGCCCATCGGCACCGTGGACGGCTTTCGCCGTA 9478
Qy 875 TTGCCAAGCGGGATACGAGATCGGGGGCACCTGATTAAAGAGGTGATATGGTTTGG 934
Db 9479 CGGCCACCGAGGAGTCTGAGCTCGGGGAGTCTTGCTCAGGCGCGGGAGACGGTCGTGC 9538
Qy 935 CGTTTGTGGCATCGSCAAATCGTGATGAAGCAAAAGTTTGAAGACCGGCACATGTTTGATA 994
Db 9539 CGTCGATGGGCGCGCCCAACCGGACCCCGAGCTGTTACGAGACCCCGACGAGCTGGACC 9598
Qy 995 TCCGCGGCCATCCCAATCCGCAATATGCGTATGCTTGGCCACCGCATCCATTTTGCCTTGGG 1054
Db 9599 TCGCGCGGCGCGCAATCCGACCTTGGGCTTGGGCGCGGACCGCACCACTGCCCTGGGCG 9658
Qy 1055 CCGCGCTTGGCCGCTCTTGAAGCAATATCGGTTAAACGTCCTTTGAT 1100
Db 9659 CCCAACTGGCCCGGCTGGAGCTCCAGATCACGCTCACGACGCTGTT 9704

Search completed: May 28, 2004, 22:10:36
Job time : 108 secs

> O <
O | O IntelliGenetics
> O <

GENALIGN - Multiple Sequence Alignment Program
Release 5.4

Wed 2 Jun 104 12:59:50-PDT

Solution Parameters:

Nucleic Alphabet = Identity
Output line length = 80
Compress = Off
Histogram = Off
Randomization = Off
AMINO-Res-length = 2
Deletion-weight = 5.00
Length-factor = 0
Matching-weight = 1.00
NUCLEIC-Res-length = 4
Spread-factor = 50

Clustered order of selected sequences:

1. US-09-869-334B-2 (1-1191)
40. US-09-869-334B-41 (1-1204)
41. US-09-869-334B-43 (1-1221)
42. US-09-869-334B-44 (1-1221)

Region Alignment: (listed in Clustered order)

```

US-09-869- 1      ATGATGTTTAAACCGCGCGCAAGCCTTGCGAGCGAG
US-09-869- 1      GTCGAC  aATGAATGTTTAAACCGCGCGCAAGCCTTGCGAGCGAG
US-09-869- 1      CTCGAGTCGAGAGTCGACTAATATGAACGTTCTGAACCGCGCGTCAAGCCTTGCGAGCGAG
US-09-869- 1      CTCGAGTCGAGAGTCGACTAATATGAACGTTCTGAACCGCGCGTCAAGCCTTGCGAGCGAG
consensus      ctcgagtcgagaggtcgactaatatgaa-gt--T-AACCGCGG-CAAGCCTTGCGAGCGAG

US-09-869- 38      CGCTGCTCAATGGGAAAAACAAACAGGATGCGTATCATCCGTTTCATGGTATGAATCGAT
US-09-869- 45      CGCTGCTCAATGGGAAAAACAAACAGGATGCGTATCATCCGTTTCATGGTATGAATCGAT
US-09-869- 62      CGCTGCTCAATGGGAAAAACAAACAGGATGCGTATCATCCGTTTCATGGTATGAATCGAT
US-09-869- 62      CGCTGCTCAATGGGAAAAACAAACAGGATGCGTATCATCCGTTTCATGGTATGAATCGAT
consensus      CGCTGCTCAATGGGAAAAACAAACAGGATGCGTATCATCCGTTTCATGGTATGAATCGAT

US-09-869- 99      GAGAAAGGATGCGCTGTTTCTTCTTGAAGAAAAACCAAGTGTGAGAGCGTTTTCCTTTAT
US-09-869- 106     GAGAAAGGATGCGCTGTTTCTTCTTGAAGAAAAACCAAGTGTGAGAGCGTTTTCCTTTAT
US-09-869- 123     GAGAAAGGATGCGCTGTTTCTTCTTGAAGAAAAACCAAGTGTGAGAGCGTTTTCCTTTAT
US-09-869- 123     GAGAAAGGATGCGCTGTTTCTTCTTGAAGAAAAACCAAGTGTGAGAGCGTTTTCCTTTAT
consensus      GAGAAAGGATGCGCTGTTTCTTCTTGAAGAAAAACCAAGTGTGAGAGCGTTTTCCTTTAT

US-09-869- 160     GATGATGCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCGAGA
US-09-869- 167     GATGATGCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCGAGA
US-09-869- 184     GATGATGCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCGAGA

```

```

US-09-869- 184  GATGATGTCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCGAG
consensus      GATGATGTCAAAAAAGTTGTTGGGATAAAGAGTTGTTTTCAGTTGCAATGCGCGAGCGAG

US-09-869- 221  CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC
US-09-869- 228  CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC
US-09-869- 245  CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC
US-09-869- 245  CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC
consensus      CAAGCTCTATTGGAATTCATTAATCATGACCCGCGGAAGCATACAAAAATCCGTTTC

US-09-869- 282  AGTCGTGAAACAAAGCCTTTACTCCGCGCG- GATGAAGCAATGGGAACCGGAATTCAGGAA
US-09-869- 289  AGTCGTGAAACAAAGCCTTTACTCCGCGCGCGGATGAAGCAATGGGAACCGGAATTCAGGAA
US-09-869- 306  AGTCGTGAAACAAAGCCTTTACTCCGCGCG- GATGAAGCAATGGGAACCGGAATTCAGGAA
US-09-869- 306  AGTCGTGAAACAAAGCCTTTACTCCGCGCG- GATGAAGCAATGGGAACCGGAATTCAGGAA
consensus      AGTCGTGAAACAAAGCCTTTACTCCGCGCG- GATGAAGCAATGGGAACCGGAATTCAGGAA

US-09-869- 343  ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGGATT
US-09-869- 350  ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGGATT
US-09-869- 367  ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGGATT
US-09-869- 367  ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGGATT
consensus      ATCACAGATGAATGATTCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTCAGGATT

US-09-869- 404  TTTTCATACCGCTTCGGTTATTGTGATATCTGAGTCTGCGGAGTSCCTTTCAGGCGAGAT
US-09-869- 411  TTTTCATACCGCTTCGGTTATTGTGATATCTGAGTCTGCGGAGTSCCTTTCAGGCGAGAT
US-09-869- 428  TTTTCATACCGCTTCGGTTATTGTGATATCTGAGTCTGCGGAGTSCCTTTCAGGCGAGAT
US-09-869- 428  aTTTCATACCGCTTCGGTTATTGTGATATCTGAGTCTGCGGAGTSCCTTTCAGGCGAGAT
consensus      TTTTCATACCGCTTCGGTTATTGTGATATCTGAGTCTGCGGAGTSCCTTTCAGGCGAGAT

US-09-869- 465  GGAACAGTTTAAAGCATGCTCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAA
US-09-869- 472  GGAACAGTTTAAAGCATGCTCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAA
US-09-869- 489  GGAACAGTTTAAAGCATGCTCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAA
US-09-869- 489  GGAACAGTTTAAAGCATGCTCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAA
consensus      GGAACAGTTTAAAGCATGCTCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAA

US-09-869- 526  GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTTG
US-09-869- 533  GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTTG
US-09-869- 550  GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTTG
US-09-869- 550  GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTTG
consensus      GCTGAAAAAGCCTTTTGGAGAACGAGATAGTGTGAGGAAGAACTGGCGCGGTTTTTTG

US-09-869- 587  CCGGCATCATAGAGAAACGAAACAAACCGGAACAGGATATTATTTCTATTATGAGTGA

```



```
US-09-869- 594 CCGCATCATAGAAAGAAAGCGAAACAAACCGGAACAGGATATATTCTATTTTAGTGGAA
|
|
|
US-09-869- 611 CCGGCATCATAGAAAGAAAGCGAAACAAACCGGAACAGGATATATTCTATTTTAGTGGAA
|
|
|
US-09-869- 611 CCGGCATCATAGAAAGAAAGCGAAACAAACCGGAACAGGATATATTCTATTTTAGTGGAA
|
|
|
consensus
|
|
|
US-09-869- 648 AGCGGAAGAAACAGCGGAGAGAGCTGTCGGGTGAAGAGCTGATTCCGTTTGCACGCTGCTG
|
|
|
US-09-869- 655 AGCGGAAGAAACAGCGGAGAGCTGTCGGGTGAAGAGCTGATTCCGTTTGCACGCTGCTG
|
|
|
US-09-869- 672 AGCGGAAGAAACAGCGGAGAGCTGTCGGGTGAAGAGCTGATTCCGTTTGCACGCTGCTG
|
|
|
US-09-869- 672 AGCGGAAGAAACAGCGGAGAGCTGTCGGGTGAAGAGCTGATTCCGTTTGCACGCTGCTG
|
|
|
consensus
|
|
|
US-09-869- 709 CTGCTGCGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAG
|
|
|
US-09-869- 716 CTGCTGCGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAG
|
|
|
US-09-869- 733 CTGCTGCGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAG
|
|
|
US-09-869- 733 CTGCTGCGCGGAAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAG
|
|
|
consensus
|
|
|
US-09-869- 770 AAACGCCAGCGGTTTACAGAGAACTGGCAGCCATCTGAACTGATGCTCAGGCACTGGGA
|
|
|
US-09-869- 777 AAACGCCAGCGGTTTACAGAGAACTGGCAGCCATCTGAACTGATGCTCAGGCACTGGGA
|
|
|
US-09-869- 794 AAACGCCAGCGGTTTACAGAGAACTGGCAGCCATCTGAACTGATGCTCAGGCACTGGGA
|
|
|
US-09-869- 794 AAACGCCAGCGGTTTACAGAGAACTGGCAGCCATCTGAACTGATGCTCAGGCACTGGGA
|
|
|
consensus
|
|
|
US-09-869- 831 GGAAGCCTTCGCTTTCAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAAGCGGGATACG
|
|
|
US-09-869- 838 GGAAGCCTTCGCTTTCAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAAGCGGGATACG
|
|
|
US-09-869- 855 GGAAGCCTTCGCTTTCAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAAGCGGGATACG
|
|
|
US-09-869- 855 GGAAGCCTTCGCTTTCAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAAGCGGGATACG
|
|
|
consensus
|
|
|
US-09-869- 892 GAGATCGGGGGGACCTGATTAAGAAGGTGATATGTTTTCGGGTTTGGCATCGGCAG
|
|
|
US-09-869- 899 GAGATCGGGGGGACCTGATTAAGAAGGTGATATGTTTTCGGGTTTGGCATCGGCAG
|
|
|
US-09-869- 916 GAGATCGGGGGGACCTGATTAAGAAGGTGATATGTTTTCGGGTTTGGCATCGGCAG
|
|
|
US-09-869- 916 GAGATCGGGGGGACCTGATTAAGAAGGTGATATGTTTTCGGGTTTGGCATCGGCAG
|
|
|
consensus
|
|
|
US-09-869- 953 ATCGTGATGAAGCAAGTTTTCAGACCGGCACATGTTTGTATATCCGCGGCCATCCCAATCC
|
|
|
US-09-869- 960 ATCGTGATGAAGCAAGTTTTCAGACCGGCACATGTTTGTATATCCGCGGCCATCCCAATCC
|
|
|
US-09-869- 977 ATCGTGATGAAGCAAGTTTTCAGACCGGCACATGTTTGTATATCCGCGGCCATCCCAATCC
|
|
|
US-09-869- 977 ATCGTGATGAAGCAAGTTTTCAGACCGGCACATGTTTGTATATCCGCGGCCATCCCAATCC
|
|
|
consensus
|
|
|
```

```
US-09-869- 1014 GCATATTGCGTTTGGCCACGCGATCCATTTTTCCTTGGGGCCCGCTTGCCTGCTTTGAA
|
|
|
US-09-869- 1021 GCATATTGCGTTTGGCCACGCGATCCATTTTTCCTTGGGGCCCGCTTGCCTGCTTTGAA
|
|
|
US-09-869- 1038 GCATATTGCGTTTGGCCACGCGATCCATTTTTCCTTGGGGCCCGCTTGCCTGCTTTGAA
|
|
|
US-09-869- 1038 GCATATTGCGTTTGGCCACGCGATCCATTTTTCCTTGGGGCCCGCTTGCCTGCTTTGAA
|
|
|
consensus
|
|
|
US-09-869- 1075 GCAAAATATCGGTTTAAACGCTCTTTGATTTCTGTTTCTCATATGAGAGTGGGTCAAGTATCA
|
|
|
US-09-869- 1082 GCAAAATATCGGTTTAAACGCTCTTTGATTTCTGTTTCTCATATGAGAGTGGGTCAAGTATCA
|
|
|
US-09-869- 1099 GCAAAATATCGGTTTAAACGCTCTTTGATTTCTGTTTCTCATATGAGAGTGGGTCAAGTATCA
|
|
|
US-09-869- 1099 GCAAAATATCGGTTTAAACGCTCTTTGATTTCTGTTTCTCATATGAGAGTGGGTCAAGTATCA
|
|
|
consensus
|
|
|
US-09-869- 1136 CTCGANTTGAACACAGTGTGATATACGGATTAAAGAGCTTCGGTGTGAAATGTAA
|
|
|
US-09-869- 1143 CTCGANTTGAACACAGTGTGATATACGGATTAAAGAGCTTCGGTGTGAAATGTAA
|
|
|
US-09-869- 1160 CTCGANTTGAACACAGTGTGATATACGGATTAAAGAGCTTCGGTGTGAAATGTAA
|
|
|
US-09-869- 1160 CTCGANTTGAACACAGTGTGATATACGGATTAAAGAGCTTCGGTGTGAAATGTAA
|
|
|
consensus
|
|
|
US-09-869- 1192
|
US-09-869- 1204 C
|
US-09-869- 1221 C
|
US-09-869- 1221 C
|
consensus
|
|
|
Alignment score = 3366.00
Scoring matrix:
```

	1	40	41	42
1				
40	1123	1035	1027	
41		1111	1107	1213
42				

> O <
O||O IntelliGenetics
> O <

GENALIGN - Multiple Sequence Alignment Program
Release 5.4

Wed 2 Jun 104 12:57:39-PDT

Solution Parameters:

Amino Alphabet = Identity
Output line length = 80
Compress = Off
Histogram = Off
Randomization = Off

AMINO-Res-length = 2
Deletion-weight = 5.00
Length-factor = 0
Matching-weight = 1.00
NUCLEIC-Res-length = 4
Spread-factor = 50

Clustered order of selected sequences:

- 1. US-09-869-334B-1 (1-396)
- 2. US-09-869-334B-42 (1-396)
- 3. US-09-869-334B-45 (1-396)

Region Alignment: (listed in Clustered order)

```
US-09-869- 1 MNVLNRRQALQALLNGKKNQDAYHPPFWYESMRKDAPVSFDEENQVMSVFLYDDVKVVG
US-09-869- 1 MNVLNRRQALQALLNGKKNQDAYHPPFWYESMRKDAPVSFDEENQVMSVFLYDDVKVVG
US-09-869- 1 MNVLNRRQALQALLNGKKNQDAYHPPFWYESMRKDAPVSFDEENQVMSVFLYDDVKVVG
US-09-869- 1 MNVLNRRQALQALLNGKKNQDAYHPPFWYESMRKDAPVSFDEENQVMSVFLYDDVKVVG
consensus MNVLNRRQALQALLNGKKNQDAYHPPFWYESMRKDAPVSFDEENQVMSVFLYDDVKVVG

US-09-869- 62 DKELFSSCMPQQTSSIGNSIIMDPPKHTKIRSVVNKAFTPRMKQWEPRIOEITDELIQ
US-09-869- 62 DKELFSSCMPQQTSSIGNSIIMDPPKHTKIRSVVNKAFTPRMKQWEPRIOEITDELIQ
US-09-869- 62 DKELFSSCMPQQTSSIGNSIIMDPPKHTKIRSVVNKAFTPRMKQWEPRIOEITDELIQ
consensus DKELFSSCMPQQTSSIGNSIIMDPPKHTKIRSVVNKAFTPRMKQWEPRIOEITDELIQ

US-09-869- 123 FQGRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQFKAWSDDLIVSTPKDKSBEAEKAFLEE
US-09-869- 123 FQGRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQFKAWSDDLIVSTPKDKSBEAEKAFLEE
US-09-869- 123 FQGRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQFKAWSDDLIVSTPKDKSBEAEKAFLEE
consensus FQGRSEFDLVHDFSYPLPVIVISLGLVPSAHMEQFKAWSDDLIVSTPKDKSBEAEKAFLEE

US-09-869- 184 RDKCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGEELIPFCTLLLVAGNETT
US-09-869- 184 RDKCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGEELIPFCTLLLVAGNETT
US-09-869- 184 RDKCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGEELIPFCTLLLVAGNETT
consensus RDKCEELAAFPAGIIEKRNKPEQDIISILVEAETGKLSGEELIPFCTLLLVAGNETT

US-09-869- 245 TNLISNAMYSILETGPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDTIEIGHLIK
US-09-869- 245 TNLISNAMYSILETGPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDTIEIGHLIK
```

```
US-09-869- 245 TNLISNAMYSILETGPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDTIEIGHLIK
consensus TNLISNAMYSILETGPGVYELRSHPELMPQAVEALRFRAPAPVLRRIAKRDTIEIGHLIK

US-09-869- 306 EGDWVLAFAVASANRDEAKFDRPHMEDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSL
US-09-869- 306 EGDWVLAFAVASANRDEAKFDRPHMEDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSL
US-09-869- 306 EGDVLAFAVASANRDEAKFDRPHMEDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSL
consensus EGDWVLAFAVASANRDEAKFDRPHMEDIRRHENPHIAFGHGHFCLGAPLARLEANIALTSL

US-09-869- 367 ISAFPHMECVSITPIENSIVYGLKSPRVKM
US-09-869- 367 ISAFPHMECVSITPIENSIVYGLKSPRVKM
US-09-869- 367 ISAFPHMECVSITPIENSIVYGLKSPRVKM
consensus ISAFPHMECVSITPIENSIVYGLKSPRVKM
```

Alignment score = 785.00

Scoring matrix:

	1	2	3
1		394	389
2			391
3			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 23:04:11 ; Search time 71 Seconds
(without alignments)
536.505 Million cell updates/sec

Title: US-09-869-334B-1

Perfect score: 2052
Sequence: 1 MNVLRALORALLGNKK.....SITPIENSVYGLKSFVRVM 396

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*

- 1: Pir1:*
- 2: Pir2:*
- 3: Pir3:*
- 4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2052	100.0	396	1 B69851	cytochrome P450 Yj
2	738	36.0	410	2 S39924	cytochrome P450meg
3	711.5	34.7	410	1 O4BS6M	cytochrome P450 10
4	674	32.8	417	1 S49051	cytochrome P450 ty
5	658	32.1	410	1 S49611	cytochrome P450 cy
6	646.5	31.5	405	1 B42606	cytochrome P450 CV
7	626	30.5	376	1 G69679	polyketide hydroxy
8	600	29.2	395	1 G69594	cytochrome P450 bi
9	581	28.3	397	1 S51594	cytochrome P450 my
10	581	28.3	434	2 C87170	probable cytochrom
11	571	27.8	418	2 G83229	cytochrome P450 PA
12	565.5	27.6	412	1 B40634	erythromycin monoo
13	564	27.5	404	2 T30231	cytochrome P450 -
14	548.5	26.7	368	1 S71328	cytochrome P450 CY
15	540	26.3	406	1 S18531	cytochrome P450 er
16	538	26.2	310	2 T44767	cytochrome P450 li
17	536.5	26.1	411	2 T36526	probable cytochrom
18	532.5	26.0	411	1 J25859	polyketide synthas
19	524	25.5	415	2 A27331	cytochrome P450 hy
20	524	25.5	415	2 B97512	probable cytochrom
21	520	25.3	428	1 A70729	cytochrome P450 Rv
22	516	25.1	402	1 A70707	cytochrome P450 Rv
23	514.5	25.1	398	1 H70807	cytochrome P450 Rv
24	499	24.3	433	1 B70677	cytochrome P450 Rv
25	497.5	24.2	444	2 F87366	cytochrome P450 fa
26	489.5	23.9	402	2 H82813	cytochrome P450-1i
27	488	23.8	414	1 E70708	cytochrome P450 Rv
28	482	23.5	406	1 A48495	linalool 8-monooxy
29	481	23.4	411	1 J24003	cytochrome P450 -

ALIGNMENTS

RESULT 1

B69851

Cytochrome P450 yj1B - Bacillus subtilis

N:Contains: oxidoreductase (BC 1.-.-.-)

C:Species: Bacillus subtilis

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #ext_change 16-Jun-2000

C:Accession: B69851

R:Kunst, P.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmeron, P.I.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Kotter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portecelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: B69851

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-396 <KJN>

A:Cross-references: GB:299110; GB:AL009126; NID:G26333472; PIDN:CAB13078.1; PID:G2633575

A:Experimental source: strain 168

C:Genetics:

A:Gene: yj1B

C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: oxidoreductase

F:236-371/Domain: cytochrome P450 homology <CYP>

Query Match 100.0%; Score 2052; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVLRALORALLGNKKQDAYHPFPWYEMRKDAPVSFDEENQVNSVFLYDDVKVY 60

DB 1 MNVLRALORALLGNKKQDAYHPFPWYEMRKDAPVSFDEENQVNSVFLYDDVKVY 60

QY 61 GDXELFSSCMPQOTSIGNSIINMDPPKTKRSVVKAFTRVVKMKEPRIOETDELI 120

DB 61 GDXELFSSCMPQOTSIGNSIINMDPPKTKRSVVKAFTRVVKMKEPRIOETDELI 120

QY 121 QKQGRSEFDLVHDFSYPLPVIIVISLGVPSAHMQFKANSDLLVSTFKKSEAEKAF 180

DB 121 QKQGRSEFDLVHDFSYPLPVIIVISLGVPSAHMQFKANSDLLVSTFKKSEAEKAF 180

QY 181 LEERDKCEELAAFPAGIIEBKRNKPEQDIISILVEAETGEKLSGELIPPTLLLVAG 240

```
Db 181 LBERKQCEELAAFGNGIIEERKNPEQDIISLVEAETGKLSGSELIPFCTLLVAG 240
Qy 241 NETTNLSNAMYSLTSGVYBELSHPELMPQAVEEALRPAPVLRRAKRTTEIG 300
Db 241 NETTNLSNAMYSLTSGVYBELSHPELMPQAVEEALRPAPVLRRAKRTTEIG 300
Qy 301 GHLIEGDMVLAFVANSANDEAKFDRPHMFDIRRHNPFIAPGHGHPCLGAPLARLEAN 360
Db 301 GHLIEGDMVLAFVANSANDEAKFDRPHMFDIRRHNPFIAPGHGHPCLGAPLARLEAN 360
Qy 361 IALTSLISAPFPMCEVSIPIENSIVYGLKSPVXM 396
Db 361 IALTSLISAPFPMCEVSIPIENSIVYGLKSPVXM 396

RESULT 2
cytochrome P450meg - Bacillus megaterium (ATCC 13368)
C:Species: Bacillus megaterium
A:Variety: ATCC 13368
C:Date: 07-Oct-1994 #sequence_revision 23-Feb-1996 #text_change 28-Jul-2000
R:Auschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boidol, W.; Siewert, G.
Mol. Gen. Genet. 241, 170-176, 1993
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the steroid
F:240-377/Domain: cytochrome P450 meg, the steroid
F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted
A:Accession: S33924
A:Molecule type: DNA
A:Residues: 1-410 <GRAU>
A:Cross-references: EMBL:221972; NID:G288298; PIDN:CAA79985.1; PID:G2838300
A:Experimental source: ATCC 13368
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homolog
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:240-377/Domain: cytochrome P450 homolog <P45>
F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted
Query Match 36.0%; Score 736; DB 2; Length 410;
Best Local Similarity 41.1%; Pred. No. 1.3e-43;
Matches 157; Conservative 74; Mismatches 133; Indels 18; Gaps 5;
Qy 18 KKKQDAYHFPWYSGMRKQAPVDFEENQVMSVFLYDDKVVGVGKELFSSCMPQQTSSI 77
Db 14 KRTBEFSYAKCKRMLNDPVSYHGTDTWNVFYEDVKVLSYKHFSSVRKRTTISV 73
Qy 78 GN-----SIINMDPPKHTKIRSVVNVKAFTRPVKQWEPRIQETDELQIQEQR 126
Db 74 GTDSREGSVPEKIQTESDPPHRRKRSLLAAFTPRSQNWEPRIQETDELQIQEQR 133
Qy 127 SEFDLVHDFSYPLPVIIVISLGLVPSAHMEQKAWSD-LLVSTPKDKSEAEKAFLEERD 185
Db 134 TEIDIVASLASPLIIVMADLGVPSKORLLPKKVVDTLFLPDRKQEVVDKL----KQ 189
Qy 186 KCEBELAFAAGIIEKRNKPEODIISLVEAETGKLSGSELIPFCTLLVAGNETTT 245
Db 190 VAAKGYQYLYPIVQKRLNPADDIISLKSVDGEMFTDDSEVATVMTLLGAGVETTS 249
Qy 246 NLISNAMYSL-ETPGVYBELSHPELMPQAVEEALRPAPVLRRAKRTTEIGGHLI 304
Db 250 HLGANSFSLLYDDKEVYQELHNDLVQAVEEMLRFRNLKLDRTVKEDNDLVGL 309
Qy 305 KEGDMVLAFVANSANDEAKFDRPHMFDIRRHNP-PIAFCHGHPCLGAPLARLEANIAL 363
Db 310 KEGDSVVVWMSAANDMEEMFDPFTLIHRRPNKHLTKGNGPHFCLGAPLARLEANIAL 369
Qy 364 TSLISAPFPMCEVSIPIENSIV 385
Db 370 TAPLKKFKHIEAVSPQLEENL 391

RESULT 3
C4586M
cytochrome P450 106 - Bacillus megaterium
```

```
N:Alternate names: cytochrome P450BM-1
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus megaterium
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Mar-2000
C:Accession: S07764; S17973
R:He, J.S.; Ruettinger, R.T.; Liu, H.M.; Fulco, A.J.
Biochim. Biophys. Acta 1009, 301-303, 1989
A:Title: Molecular cloning, coding nucleotides and the deduced amino acid sequence of P-4
A:Reference number: S07764; MUID:90089408; PMID:2597681
A:Accession: S07764
A:Molecule type: DNA
A:Residues: 1-410 <HEJ1>
A:Cross-references: EMBL:X16610; NID:G39626; PIDN:CAA34612.1; PID:G39627
A:Accession: S17973
A:Molecule type: protein
A:Residues: 1-25 <HEJ2>
C:Genetics:
A:Gene: CYP106
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homolog
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;
F:1-410/Product: cytochrome P450 106 #status experimental <MAT>
F:241-378/Domain: cytochrome P450 homolog <CYP>
F:356/Binding site: heme iron (Cys) (axial ligand) #status predicted
Query Match 34.7%; Score 711.5; DB 1; Length 410;
Best Local Similarity 39.7%; Pred. No. 9.1e-42;
Matches 151; Conservative 74; Mismatches 132; Indels 23; Gaps 6;
Qy 18 KKKQDAYHFPWYSGMRKQAPVDFEENQVMSVFLYDDKVVGVGKELFSSCMPQQTSSI 77
Db 15 QSRAEFFPIQWYKEMNNNSPVYFHEETNTWNVFYEHVKQVLSNYDFFSSDQRTTIFV 74
Qy 78 GN-----SIINMDPPKHTKIRSVVNVKAFTRPVKQWEPRIQETDELQIQEQR 126
Db 75 GDSKKKSTSPITNLNLPDPRKARSLLAAFTPRSLKQWEPRIKQIAADLVEALQKN 134
Qy 127 SEFDLVHDFSYPLPVIIVISLGLVPSAHMEQKAWSDLLVST-PKDKSEAEKAFLEERD 185
Db 135 STINIVDDLSPPSLVIADLFGVPVKDRYQFKKVVDTLFPQVDOERLEIE---CEKQ 190
Qy 186 KCEBELAFAAGIIEKRNKPEODIISLVEAETGKLSGSELIPFCTLLVAGNETTT 245
Db 191 RAGAEYQYLYPIVIEKRNLSODIISDLIQAEVDGETTDEIVHATMLLGAGVETTS 250
Qy 246 NLISNAMYSL-ETPGVYBELSHPELMPQAVEEALRPAPVLRRAKRTTEIGGHLI 304
Db 251 HALANFYSLYDKSLSELNRELAKPAVEEMLRYRHFHISRDRTVKQDNELLGVKL 310
Qy 305 KEGDMVLAFVANSANDEAKFDRPHMFDIRRHNP-PIAFCHGHPCLGAPLARLEANIAL 363
Db 311 KKGDDVVVWMSAANDMEEMFDPFTLIHRRPNKHLTKGNGPHFCLGAPLARLEANIAL 370
Qy 364 TSLISAPFPMCEVSIPIEN 383
Db 371 EAPLEAFSHIE-----PFED 385

RESULT 4
S49051
cytochrome P450 tyli (EC 1.1.-.-) [validated] - Streptomyces fradiae (strain T59235)
C:Species: Streptomyces fradiae
A:Variety: strain T59235
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S49051
R:Merison-Davies, L.A.; Cundliffe, E.
Mol. Microbiol. 13, 349-355, 1994
A:Title: Analysis of five tylosin biosynthetic genes from the tyliBA region of the Strept
A:Reference number: S49051; MUID:95075319; PMID:7984112
A:Accession: S49051
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-417 <MER>
A:Cross-references: EMBL:U08223; NID:G6849140; PIDN:AAA21341.1; PID:G473597
```

A;Status: translated from GB/EMBL/DBCU
A;Molecule type: DNA
A;Residues: 1-410 <BEL>
A;Cross-references: EMBL:Y11043; NID:g1326275; PIDN:CAA71937.1; PID:g1326278
A;Experimental source: strain 1A1
C;Genetics:
A;Gene: cypA
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;245-381/Domain: cytochrome P450 homology <CYP>

Query Match 32.1%; Score 658; DB 1; Length 410;
Best Local Similarity 37.4%; Pred. No. 4.6e-38;
Matches 146; Conservative 78; Mismatches 136; Indels 30; Gaps

QY 1 MNVINRQALQRALLGNKKKQDAYHFFEFVYESMKDA--PVSFDDEENQVMSVELYDDVK 5
:
Db 9 VTILTESQLSSRAF-----KDEAYE--FYKELRKSQALYELSGALGKGWLSRYDDAI 6

QY 58 KVVGDKEI-----FSS-----CMPQQTSISIGNSIINMDPPKHKTIRSVYNKAFTPRVM 1
:
Db 61 HLLGNKELKKKNVENVFATAKEKRPAALLKNEETLTTHMLNSDPDHNRLTIVIQKAFTRMI 1

QY 106 KWPEPIQEITDELIOKFQGRSEFDVHDPSYPLPVIVISELLGVPSAHWEQFAKWSDLL 1
:
Db 121 LQLEDKIQLADSLLDKVPNKFWNLVCEYAPFLPIIVISEMLGIPLEDKQKFRVWSQAI 1

QY 166 VSTPKDKSEAEAKALEERDKCEEELAAFPAGIIIEKRNKPEDDIISILVEAETGEKLKS 2
:
Db 181 I----DFSDAPER--LQENDHLLEGFEVYLESTRVKRRREPAGSLIASLIQAESegTQLS 2

QY 226 GEELIFFCTLLVAGNETTNLTLSNAMYSLTEPGVVEELRSHPELMPQAVEALRPAP 2
:
Db 235 TEELYSMILLIVAGHEITVNLTNTYALMHCHHQLEKURQOPDLMNSAIEEARLFHSP 2

QY 286 APVLR-RIAKRDTIEGGHLIKEGDMVLAFVASANRDEAKEPDHFMDIREHNPPIAFGH 3
:
Db 295 VELTTIRWTAEPIFHGQEIKKRDVIIISLASNRDEKIFPNADIFDERKNNRHIAFGH 3

QY 345 GHFCIGCAPLARLENIALTSLISAFPHME 374
:
Db 355 GNHFICGAOLARLEAKIAISTLLERCENIQ 384

RESULT 6
B42606
CYTOCHROME P450 CYP106 - Saccharopolyspora erythraea
N;Contains: oxidoreductase (EC 1.-.-.)
C;Species: Saccharopolyspora erythraea
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: B42606
R;Andersen, J.F.; Hutchinson, C.R.
J. Bacteriol 174, 725-735, 1992
A;Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 gene
A;Reference number: A42606; UID:92121109; PMID:1732208
A;Accession: B42606
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA; protein
A;Residues: 1-405 <ND>
A;Cross-references: GB:M83110; NID:g152682; PIDN:AAA26483.1; PID:g152684
A;Experimental source: NRRL2338
A;Note: Sequence extracted from NCBI backbone (NCBIP:77484)
C;Genetics:
A;Gene: CYP107B1
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F;238-374/Domain: Cytochrome P450 homology <CYP>
F;352/Binding site: heme iron (Cys) (axial ligand); #status predicted

Query Match 31.5%; Score 646.5; DB 1; Length 405;
Best Local Similarity 37.8%; Pred. No. 2.8e-37;
Matches 138; Conservative 72; Mismatches 122; Indels 33; Gaps

QY	30	YESMRKDAPVSGDEEN--QVWSVFLYDDVKKVGVGKELP-----SSCMP 71
Db	24	YARMREBPQVORITVNGLDLAWLITRYEDVQKALLDPRJAKFGRTQOIIEKSLDAERP 83
QY	72	QQTSSIGNSIINMDPPKHTKIRSVVNKAFTPRVNMKWEPRIQEIIDELIQFGSGSEDL 131
Db	84	GFSGDLGPHMLNTPPDHTRUKLVVKAFTARVVEGLRPRIEQITDDLLDRLAGSEVDL 143
QY	132	VHDFSYELPVIIVISELLGVPSSAHMBQKAWSDLLVSTPKDKE--BAEKAFLEERDKCEE 190
Db	144	IDFEAFPLPTIVISELMGVDSRRDDPSWINVLVDSQPEAQASVAMVE----- 195
QY	191	LAFFAGIIEEKRNKPPQDITISILVERBETGEKUSGEBELIPCTLLVAGNETTNLTSN 250
Db	196	---YLTELIAKRTPEPGDILLTALLEAVEVDGORTSEGLIAMVFLLVAGHETTNNLIGN 252
QY	251	AMYSILETPGVYELSRHPLEMPQAVEALSRAPAP--VLIRRIAKRTEIGHLTKGCD 308
Db	253	CVLSSLGNPQOLAALRNDPSLLPGAIETLRYESPVANGTFRHTAEA--VRFGDVWIPEGE 311
QY	309	MVLAFVASANRDEAKFRDPRHMFDRRHNPHIAFGHGHFCIGAPLARLEANNIALTSIFS 368
Db	312	LVMVALGAANRDGERFEDPDRFDITREITGHVAFGEGIHFCVCGAALANLEAQIANGRELLE 371
QY	369	AFPHM 373
Db	372	RFPOL 376

RESULT 7
G63679
polyketide hydroxylase (EC 1.-.-.-) pkss - *Bacillus subtilis*
C:Species: *Bacillus subtilis*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: G69679
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C:Bron, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koster, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, C.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MJID:98044033; PMID:9384377
A:Accession: G69679
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-376 <KUN>
A:Cross-references: GB:Z39113; GB:ALC09126; NID:G2634090; PIDN:CAB13607.1; PID:G2634107
A:Experimental source: strain 168

```

Query Match      30.5%  Score 626;  DB 1;  Length 376;
Best Local Similarity 35.4%  Pred. No. 6.7e-36;
Matches 132;  Conservative 78;  Mismatches 133;  Indels 28;  Gaps 4
QY      10  LQALLNGKQDAYHFPFVYESMRKQAPV-----SPDEENQVWSVLYD----- 54
      : : : : : : : : : : : : : : : : : : : : : : : :
Db       3  MEXLMFPHGKEFHNNPFSVLGRFREPEEPZHRFELKRFGATVPALITRYDDCMFLKDN 62
      : : : : : : : : : : : : : : : : : : : : : : : :
QY      55  ----DVKKVYGVQDKELFSSCMPQQTSSIGNSINMDPPNKHTRKSVVNKAFTPRVKKQWEP 110
      : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

RESULT 8

G69594
cytochrome P450 bioI - Bacillus subtilis
N:Contains: oxidoreductase (EC 1.-.-.)
C:Species: Bacillus subtilis
C>Date: 10-Sep-1993 #sequence_revision 10-Sep-1993 #text_change 16-Jun-2000
C:Accession: G69594
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Aloni, G.; Azevedo, V.; Berterrotich, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hulio, M.F.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parto, V.; Poh, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, E.; Sekiguchi, J.; Sekowska, A.; Seror, teuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrtra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yanamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshikawa, K.A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G69594
A>Status: nucleic acid sequence not shown; translation nct shown
A:Molecule type: DNA
A:Residues: 1-395 <KUN>
A:Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CABI4997.1; PID:g2635503
A:Experimental source: strain 168
C:Genetics:
A:Gene: bioI
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
E:z32-z367/Domain: Cytochrome P450 homology <CYP>

Query Match	29.2%	Score 600;	DB 1;	Length 395;
Best Local Similarity	35.5%;	Pred. No. 4.5e-34;		
Matches	128; Conservative	84; Mismatches	133; Indels	7;
Gaps				
QY	25	HPEFWYESMRKDAPV---	SDEENQVWSVFLLDDVKVVGDKEL-----	FSSCMPQQT 74
Dd		::::: :: :	:: :: :: :	::
Dd	15	NPYSFYDTRLAVHPVYKGSP-	LKYPGWVTVCYEETAAILKDARFKVGTPLPESSTKYQDL 73	
		:	:	:
QY	75	SSIGNS-IINMDPKHKIRSVNKAFTPRVMKOWEPIQEITDELIOKFQGRSEFDLVH 133		
Dd		::::: :: :	:: :: :: :	::
Dd	74	SHVQNQMFLFQNGPDHRRLRTLAGATGTPRTSYQVPYIIETVVHLLDVOVQKKKNNEVIS 133		
		:	:	:
QY	134	DFSYPFLVIVISELLGVPSAHMEQFAWSDLLVST-PKDGEEAEAKAFLEERDKCEELA 192		
Dd		::::: :: :	:: :: :: :	::
Dd	134	DRAPPLASFIYANTIGYPEEDRELKEWAASLIOTIDETRKRKA---LTEGNIMAVOAM 189		
		:	:	:


```

Query Match      27.8%; Score 571; DB 2; Length 418;
Best Local Similarity 34.1%; Pred. No. 4.9e-32;
Matches 124; Conservative 72; Mismatches 138; Indels 30; Gaps 6;

QY 30 YESMRKDA--PVSFDEENQVSVFLDYDDVKVKVGD-----KELFSMCMPOQTSS 76
DB 20 YDGLQVDGAPRAHRAAEHPVWVTVYDAEKVLNHPQVRDAPQAAELYAKRTGSPAG 79
QY 77 IGNSI-----INMOPPKTKTIRSVNVKQAFTRPVKMQWEPRTQEIITDELIQKPGRSEPDLY 132
DB 80 IGEGLSHMLNLDPDPDTRILRSLVGRAFTPRQVERLQPHIERITTEALLDAMAGREQADLM 139
QY 133 HDPSYPLPVTVISLGLGVPSSAHMQPK-AKSDLLVSTPKDSEAEAKAFLEERDKCHEEL 191
DB 140 ACPAIPUTTAVIFELGIGPAEREHAKQSM-----ERQALLSPPEAQALDAQ 188
QY 192 AAFPAGIIEBKRNKPEODIISILVEASETEGKLSGEBLIPFCTILLVAGNETTNNLSNA 251
DB 189 VDYLRLVLEAKRQPADVDVYSLGVQAADSGSLSEAEVLMSMAHLMMSGFETTNMIGNA 248
QY 252 MYSLITPGVYBELRSHPELMPQAVEALRFRAPAV-LRRIAKRDTTEIGGHLIKGDMV 310
DB 249 LVTLVNPPEQLAKLRQPELLPNAMELVRHDSFVRASMLRFTVEDVELDGVVTIPAGEYI 308
QY 311 LAFVASANRDEAKFDRPHMFDIRRRPNPHIAGHGHIHFLGALPLARLEANIATLSLSAF 370
DB 309 LVSNLTANHDAERFDDEDRDLDTENTDGHLYGFGVHYCVGASLARLEGRIATQRLIARP 368
QY 371 PNMK 374
DB 369 PDQ 372

```

```

RESULT 12
B40634
erythromycin monooxygenase (EC 1.14.-.-) - Saccharopolyspora erythraea
N:Alternate names: erythromycin C-12 hydroxylase EryK
C:Species: Saccharopolyspora erythraea
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C:Accession: B40634
R:Stassi, D.; Donadio, S.; Staver, M.J.; Katz, L.
J. Bacteriol. 175, 182-189, 1993
A:Title: Identification of a Saccharopolyspora erythraea gene required for the final hyd
A:Reference number: A40634; UID:93106953; PMID:8416893
A:Accession: B40634
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <STA>
A:Cross-references: GB:U82823; GS:L05776; NID:g2327012
A:Note: sequence extracted from NCBI backbone (NCBIN:121243, NCBI:P121245)
C:Genetics:
A:Gene: CYP113A1
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:238-376/Domain: cytochrome P450 homology <CP>
F:354/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match          27.6%; Score 565.5; DB 1; Length 412;
Best Local Similarity 39.1%; Pred. No. 1.2e-31;
Matches 136; Conservative 50; Mismatches 151; Indels 11; Gaps 6;

QY      29 WYESMRKDAPSVFDSEBNQVMSVFLYDDVKXVGCDKELFS---CMFQQTSSIGNSTINMD 85
        | :|:|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Ddb     33 WLGTWREKQP-V-WQDRYGVMHVRHADVTGLRDATFSDPTRVEGASPTPGMIIHEID 91
        || :|:|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||

QY      86 PPHTKIRSVNKAFTRVMKQWEPRIQBITDELTIQFGSRSEFOLVHDPSPLPVIVIS 145
        || :|:|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Ddb     92 PPEHRALRKWSASFTPTISDLPRIRDVTRSLLA--DAGESFOLVDVLAFPPLVTIVA 149
        || :|:|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||

QY      146 ELLGVPASHMEGFKAWSOLLVSTPKDKSEAKAFLEERDKCEELAAPFAGIIEKNRK 205
        || :|:|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Ddb     150 ELGLGPPMDHEIFGDMWSGALVDIQMD--DPTDPALAERIADVINFPTYIKARCAERRAD 207
        || :|:|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||

```

QY 206 PEQDIISILVAAEETGKLSGEELIPCTCLLLVAGNETTNLLISNAMYSLLETGPGVYEL 267

Db 208 PGDDLSRLVLAEDWGSRALDDEEAANFSTALLAGHITTTTLLGLNIVTLDDEHPAHWDA 267

QY 266 RSHPELMPQAVEELRFPAPAPVLRIAKRDTEIGGHLIKEGDMVLAFVASANRDEAKFD 325

Db 268 AEDPARIPAIVEEVLRYRPPPPQMORTTKATEVAGVPIPADVMVNTVWLSANRSDAHD 327

QY 326 RPHMED--IRRHKNPHI-AFGHGTHFCGLGAPLARLEANIALTSLISAF 370

Db 328 DPDFDPSAQVRPAPRTSSFGHGVHFCCLAPLARLENKVALEELIAR 375

RESULT 13

T30231

Cytochrome P450 - Streptomyces hygroscopicus

C/Species: Streptomyces hygroscopicus

C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 16-Feb-2001

C/Accession: T30231

K/Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L. Gene 169, 9-16, 1996

A/Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus

A/Reference number: Z20782; MUID:96186894; PMID:8635756

A/Accession: T30231

A>Status: preliminary; translated from GE/EXBL/DBJ

A/Molecule type: DNA

A/Residues: 1-404 <RPA>

A/cross-references: EMBL:X86780; NID:g987088; PTD:g987105; PIDN:CAA60465.1

C/Genetics:

A/Gene: rapN

C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology F/238-375/Docain: cytochrome P450 homology <P45>

[illegible]

C:Accession: S71328
R:Wright, R.L.; Harris, K.; Solow, B.; White, R.H.; Kennelly, P.J.
FBS Lett. 384, 235-239, 1996
A:Title: Cloning of a potential cytochrome P450 from the Archaeon Sulfolobus solfataricus
A:Reference number: S71328; MUID:96197795; PMID:8617361
A:Accession: S71328
A:Molecule type: DNA
A:Residues: 1-368 <WRI>
A:Cross-references: EMBL:U51337; NID:g1256447; PIDN:AAB03278.1; PID:g1256448
A:Experimental source: ATCC 35091
C:Genetics:
A:Gene: CYP119
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:206-339/Domain: cytochrome P450 homology <CYP>
F:317/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 26.7%; Score 548.5; DB 1; Length 368;
Best Local Similarity 34.7%; Pred. No. 1.5e-30;
Matches 134; Conservative 68; Mismatches 145; Indels 39; Gaps 10;

QY 27 PWYSEMRKDAFVSDEENQVMSVFLYDDVKKVVGDKELFSSCMP---QQTSSIGN--- 79
Db 2 YDFSEMRKKDPVYD--GNMQVFSYRYTKEVLNNFSKFSDDLTCYHRELEDLRNGKIR 59

QY 80 -----SLINKDPKXHTKIRSVNKAFTPRVWKQEPRIORITDELIOKFOQRSEFDLV 132
Db 60 FDIPTRYMLTSDPHDELRSMSADIFSPQKLTLETFTRETTRSLSDSIDPR-EDDIV 118

QY 133 HDPSVPLPVIVISSELLGVPSAHMQEKAWSDDLVSFPGKSE--EAKAFLEERDKCEE 190
Db 119 KKLAVPLPIVILSKILGLPIEKEKEKENSDDLVAFLGPGGEIFELGKKYLE----- 170

QY 191 LAAPFAGHIEEKRNKPEQDIISLVEAETGKLSGEELIPFOTLLVAGNETTNNLSN 250
Db 171 ---LIGYVKDHLNS-GTEWSRVNVS-----NLSIEKLGVIILLIAGNETTNNLSN 220

QY 251 AMYSILETPGVVEELRSHPELMPQAVEALRPRAPVLRRAKRTDTEIGGHLIKEGDMV 310
Db 221 SVIDTTRF- NLQRIEE- NLYLKAIEEALRSPVVMRTVKERVKLGQDQTEEGEYV 278

QY 311 LAVASANDKAFORPHMFDIRRHNPFIAGHGHHFCLGAPLARLEANIALTSLISAF 370
Db 279 RWIASANDREWFHDGEKFIQDRNPNLHSGSHLCLGAPLARLEARIATEEFSKRF 338

QY 371 PMECVSITPIENSIVYGLKSPRVKM 396
Db 339 RHIEILDTEKVENEVNGYKRLVRL 364

RESULT 15
S18531
cytochrome P450 eryF - Saccharopolyspora erythraea
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Saccharopolyspora erythraea
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S18531; S16745
R:Haydock, S.F.; Dowson, J.A.; Dhillion, N.; Roberts, G.A.; Cortes, J.; Leadlay, P.F.
Mol. Gen. Genet. 230, 120-128, 1991
A:Title: Cloning and sequence analysis of genes involved in erythromycin biosynthesis in methyltransferases.
A:Reference number: S18530; MUID:92079886; PMID:1840640
A:Accession: S18531
A:Molecule type: DNA
A:Residues: 1-406 <HAY>
A:Cross-references: EMBL:X60379; NID:g48941; PIDN:CAA42927.1; PID:g48943
A:Note: the authors translated the codon AGG for residue 190 as Iys
C:Genetics:
A:Gene: eryF
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
F:238-375/Domain: cytochrome P450 homology <CYP>

Query Match 26.3%; Score 540; DB 1; Length 436;
Best Local Similarity 33.7%; Pred. No. 6.6e-30;
Matches 130; Conservative 72; Mismatches 132; Indels 52; Gaps 13;

QY 20 QDAYHPPFPYSES---MRKDAPYS---FDEENQVMSVFLYDDVKKVVGDKELFSSCMPQQ 73
Db 8 ESUSFH-VDWRYRYAELRETAPVTVPVFLQDP-AWLVTYGYDEAKAALSDRLSSD--PKK 63

QY 74 -----TSSIGNSIINMDPPKHTKIRSVNKAFTPRVWKQEP 110
Db 64 KYPGVEVERPVALGFEPEDEVNVPATNMGTS---DPPTHRLRLKLVSQEFTVRVEARMP 119

QY 111 RIDEITDELIOKFOQRSEFDLVHDFSYPLPVIVISSELLGVPSAHMQEKAWSDDLVSF 169
Db 120 RVEQITAEELDDVGDGVVDIVDRFAHPLPIKVICELLGVDKEYRGFEFGWSSSILVMDP 179

QY 170 KDKSEEAFALEERDKCEEELAAFFAGITEEKENKPEQDIISLVEAETGE-KLSGE 228
Db 180 -----ERA--EQRGAAREVNFILDLVERRTTEPGDLSALIRVQDDDDGRLSADE 230

QY 229 LIFFCTLLVAGNETTNNLSNAMYSILETGTVVEEL---RSHPELMPQAVEALRFRAP 285
Db 231 LTSIALVLLLAGFEASVSLIGIGTYILLTHPD-QDQALAVRRDPSALPNAVEELRYIAP 289

QY 286 APVLRRIAKRDTTEIGGHLIKEGDMVLAFVASANDKAFORPHMFDIRRHNPFIAGHG 345
Db 290 PETTTFAAEVEEINGVAIPQYSTVLVANGAANDKQFPDPHFRFDVTRDTRGHLSFGQG 349

QY 346 IHFCLGAPLARLEANIALTSLISAF 371
Db 350 IHFCMRPLAKLEGEVALRALFORFP 375

Search completed: May 28, 2004, 23:23:41
Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 22:10:41 ; Search time 41 Seconds
(without alignments)
502.921 Million cell updates/sec

Title: US-09-869-334B-1

Perfect score: 2052
Sequence: 1 MYVLRNRALORALLGNKX.....SITPIENSVIYGLKSRVVK 396

Scoring table: BLASTM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2052	100.0	396	1 YJTB_BACSU	O34374 bacillus su
2	839	40.9	405	1 CPXW_BACSU	P27632 bacillus su
3	738	36.0	410	1 CPXW_BACME	Q06069 bacillus me
4	711.5	34.7	410	1 CPXI_BACSU	P14762 bacillus me
5	658	32.1	410	1 CPXI_BACME	O08469 bacillus su
6	646.5	31.5	405	1 CPXK_SACER	P33271 saccharopol
7	600	29.2	395	1 B101_BACSU	P53554 bacillus su
8	571	27.8	397	1 CPXQ_SACER	P48335 saccharopol
9	557	27.1	404	1 CPXJ_SACER	Q00441 saccharopol
10	548.5	26.7	368	1 CPXW_SULSO	Q55080 sulfolobus
11	520	25.3	428	1 C124_MYCTU	Q50696 mycobacteri
12	516	25.1	402	1 C123_MYCTU	P77902 mycobacteri
13	514.5	25.1	398	1 C142_MYCTU	O53563 mycobacteri
14	499	24.3	433	1 C125_MYCTU	P71856 mycobacteri
15	489.5	23.9	402	1 C13B_XYLFA	Q93955 xyliella fas
16	488	23.8	414	1 C126_MYCTU	P77903 mycobacteri
17	486.5	23.7	402	1 C13B_XYLFT	Q87ax5 xyliella fas
18	482	23.5	406	1 CPXQ_PSEPU	Q59723 pseudomonas
19	475	23.1	405	1 CPXG_STRGO	P18326 streptomyce
20	474	23.1	400	1 C13C_XYLFA	Q93966 xyliella fas
21	471.5	23.0	438	1 C140_MYCTU	Q08464 mycobacteri
22	465	22.7	399	1 C13C_XYLFT	Q87av9 xyliella fas
23	462	22.5	402	1 CPXG_STRGO	P18327 streptomyce
24	451	22.0	405	1 C130_MYCTU	Q11062 mycobacteri
25	446	21.7	381	1 CPXG_STRSQ	P23296 streptomyce
26	439	21.4	409	1 CP52_STRCC	Q59831 streptomyce
27	437.5	21.3	402	1 NOR_FUSOX	P22295 fusarium ox
28	425	20.7	399	1 FAS1_RHOFA	P26911 rhodococcus
29	424	20.7	412	1 CPXH_STRGR	P26911 streptomyce
30	419.5	20.4	428	1 CPXL_PSESP	P33006 pseudomonas
31	418.5	20.4	405	1 CPXK_BACSU	O34926 bacillus su
32	418.5	20.4	406	1 CSC4_AMYOR	Q8rn03 amycolatops
33	415	20.2	408	1 NOR2_CYLTO	Q12599 cylindrocar

34	403	19.6	403	1	NOR1_CYLTO	Q00616 cylindrocar
35	401.5	19.6	398	1	C5B3_AMYOR	Q8rn04 amycolatops
36	397	19.3	400	1	CPXP_RHISN	P55544 rhizobium s
37	391	19.1	401	1	CPXP_BRAJA	Q59203 bradyrhizob
38	390	19.0	436	1	THCB_RHOER	P43492 rhodococcus
39	395	18.8	489	1	C128_MYCTU	O59572 mycobacteri
40	376	18.3	400	1	C141_MYCTU	O28362 mycobacteri
41	375.5	18.3	429	1	CPXK_BRAJA	Q59204 bradyrhizob
42	366.5	17.9	414	1	Y4VG_RHISN	O53215 rhizobium s
43	362	17.6	414	1	CPXA_PSEPU	P00183 pseudomonas
44	355	17.3	391	1	C5A3_AMYOR	Q8rn05 amycolatops
45	351	17.1	422	1	CPXC_AGSTU	P24466 agrobacteri

ALIGNMENTS

RESULT 1
YJTB_BACSU
ID YJTB_BACSU STANDARD; PRT; 396 AA.
AC O34374;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative cytochrome P450 YJTB (EC 1.14.-.-).
GN YJTB OR BSU12210.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Rivolta C., Soldo B., Lazarevic V., Joris B., Maue C., Karamata D.;
RT "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a putative 12.3 kb operon involved in hexuronate catabolism and a perfect catabolite-responsive element".
RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst P., Ogawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Boletini A., Borchert S., Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.C., Conneron I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Entian K.D., Ewington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasabara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan B., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tarakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis".
RL Nature 390:249-256 (1997).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF015825; AAC46317.1; --
CC DR EMBL; Z99110; CAB13078.1; --
CC DR PIR; B69851; B69851.
CC DR HSP; Q55080; I109.
CC DR Subtilisin; BG13195; YjiB.
CC DR Pfam; PF00067; p450; 2.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
CC FT METAL 349 349 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SQ SEQUENCE 396 AA; 44990 MW; 9A89CF12613DBCFB CRC64;
Query Match 100.0%; Score 2052; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.6e-124; Indels 0; Gaps 0;
Matches 396; Conservative 0; Mismatches 0;
QY : MNVLRNQALQRALLGKKNQDAYHPPWYSEMRKDAVPSFDEENQVMSVFLYDDVKVW 60
DB 1 MNVLRNQALQRALLGKKNQDAYHPPWYSEMRKDAVPSFDEENQVMSVFLYDDVKVW 60
QY 61 GKELFSSCMPQOTSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKOWEPRIOBITDELI 120
DB 61 GKELFSSCMPQOTSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKOWEPRIOBITDELI 120
QY 121 QXFOGSEFPLVHDFSVPLPVIIVISLGLVPSAHMEQFKAWSDLNVSTPKDKSEAEKAF 180
DB 121 QXFOGSEFPLVHDFSVPLPVIIVISLGLVPSAHMEQFKAWSDLNVSTPKDKSEAEKAF 180
QY 181 LEERDKCEELAAFFAGIIEEKNKPEQDIISILVEAEETGKLSGSELLPFCTLLLVAG 240
DB 181 LEERDKCEELAAFFAGIIEEKNKPEQDIISILVEAEETGKLSGSELLPFCTLLLVAG 240
QY 241 NETTTLNISAMYSILETPGVYSELSHPELMPQAVEALRFPAPVLRRIAKRDTGIG 300
DB 241 NETTTLNISAMYSILETPGVYSELSHPELMPQAVEALRFPAPVLRRIAKRDTGIG 300
QY 301 GHLIKGDMVLAFVASANRDEAKFDRPHMFDIRRHNPHTAPGHGIFCLGAPLARLEAN 360
DB 301 GHLIKGDMVLAFVASANRDEAKFDRPHMFDIRRHNPHTAPGHGIFCLGAPLARLEAN 360
QY 361 IALTSLISAPPHMECVSITPIENSVIYGLKSFVRKM 396
DB 361 IALTSLISAPPHMECVSITPIENSVIYGLKSFVRKM 396
RESULT 2
CPXM BACSU STANDARD; PRT; 405 AA.
AC P27632;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome p450 109 [EC 1.14.-.-] (ORF405).
GN CYP109.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W23;
RX MEDLINE=91192601; PubMed=1849493;
RA Ahn K.S., Wake R.G.;

```

```

RT -----
RT "Variations and coding features of the sequence spanning the
RL replication terminus of Bacillus subtilis 168 and W23 chromosomes.";
RL Gene 98:107-112(1991).
CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC monooxygenases. They oxidize a variety of structurally unrelated
CC compounds, including steroids, fatty acids, and xenobiotics.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24523; AAA22720.1; --
CC DR HSP; Q55080; I107.
CC DR InterPro; IPR001128; Cytochrome_P450.
CC DR Pfam; PF00067; P450; 1.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC KW Oxidoreductase; Monooxygenase; Electron transport; Heme.
CC FT METAL 351 351 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SQ SEQUENCE 405 AA; 45845 MW; I331D5BEA74B3C04 CRC64;
Query Match 40.9%; Score 839; DB 1; Length 405;
Best Local Similarity 43.7%; Pred. No. 1e-46;
Matches 167; Conservative 90; Mismatches 117; Indels 8; Gaps 2;
QY 19 NKQDAYHPPWYSEMRKDAVPSFDEENQVMSVFLYDDVKVVGDKELFSSCMPQOTSIG 78
DB 25 SEKDLFFPPFIYDKLRRESVRYDPLRDCWDVFKYDDVQFVLKNPKLFSSKRGIT 80
QY 79 NSIINMDPPKHTKIRSVVNKAFTPRVMKOWEPRIOBITDELIQFOGSEFPLVHDFSY 138
DB 81 ESILTMPPKHTKIRSVVNKAFTPRVMKOWEPRIOBITDELIQFOGSEFPLVHDFSY 140
QY 139 LPVIVISELGLVPSAHMEQFKAWSDLNVSTPKDKSEAEKAFLEERDKCEELAAFFAGI 198
DB 141 LPVIVISELGLVPSAHMEQFKAWSDLNVSTPKDKSEAEKAFLEERDKCEELAAFFAGI 200
QY 199 IEKRNKPEQDIISILVEAEETGKLSGSELLPFCTLLLVAGNETTTLNISAMYSILET 258
DB 201 LSKERAEPKEDLMTLLQAHIDGELYLTLEQIGFCILLVAGNETTTLNISAMYSILET 260
QY 259 PGVYELRSHPELMPQAVEALRFPAPVLRRIAKRDTGIGHIFCLGAPLARLEAN 318
DB 261 SVYQQVQRQNTDANVVIETLRYSPVQAGRVATEDTELGGVFIKKGSSVISWISAN 320
QY 319 RDEAKFDRPHMFDIRRHNPHTAPGHGIFCLGAPLARLEANALTSLSAPPHMECVS- 377
DB 321 RDEKFCCKPCFKIDRPSYPHLSFGFGIHFCIGAPLARLEANALTSLSAPPHMECVS- 380
QY 378 ---ITPIENSVIYGLKSFVRKM 396
DB 381 DEKLEAIPSPFVFGVRLPVRI 402
RESULT 3
CPXM BACME
ID CPXM_BACME STANDARD; PRT; 410 AA.
AC Q06069;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome p450 (MEG) [EC 1.14.99.-] (Steroid 15-beta-hydroxylase)
DE (Steroid 15-beta-monooxygenase).
GN CYP106A2.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]

```


QY 127 SEFDLVHDSYPLPVIVISBELGVPSAHMQFKWASDLNVST-PKOKSEAEAKFALEERD 185
 Db 135 STINIVDOLSPFPFSLVIADLFGVPVDRYQFKKKWDLFPQYDQERLEETE-----QEKQ 190
 QY 186 KCEELAAAPAGILBEKRNKPEODIISILVNEBETGKLSGSELIPCTCLLVAGNETTT 245
 Db 191 RAGAEFYQDYPIVIEKRNLSDDIISDLQAEVDGTFDDEIWHATMLLGGAGVETTS 250
 QY 246 NLISNAMYSL-BTPGVYEEELRSHPELMPQAVEALFRAPAPVLRRIAKRDTGIGGHLI 304
 Db 251 HAIANFYSLYDDKLSYSELNRNRELAPRAVEMLRYRHSRDRKTVQDNELLGVKL 310
 QY 305 KEGDMVLAFVANSRDEAKTDRGEMFDIRHPN-PHIAFGHGHFCLGAPLARLENIAL 363
 Db 311 KKGDDVVIAMNSACNMDETWFENPFSVDIHRPTNKKHLTFNGNPHFCLGAPLARLEMKIIL 370
 QY 364 TSLISAPPHMECVSITPIEN 383
 Db 371 EAELEAFSHIE-----PFED 385

RESULT 5
 CPXY BACSU STANDARD; PRT; 410 AA.
 AC O08469;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome P450 (BC 1.14.-.-).
 DE CYPA OR CYP107J1 OR BSU26740.
 OS Bacillus subtilis
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=97431495; PubMed=9287000;
 RA Belitsky B.R., Gustafsson M.C.U., Sonenshein A.L., von Wachenfeldt C.;
 RT "An lrp-like gene of Bacillus subtilis involved in branched-chain
 amino acid transport."
 RL J. Bacteriol. 179:5448-5457(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=97453479; PubMed=9308178;
 RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,
 RA Duesterhoeft A., Ehrlich S.D.;
 RT "Sequence of the Bacillus subtilis genome region in the vicinity of
 the lev operon reveals two new extracytoplasmic function RNA
 polymerase sigma factors, Sigv and Sigz."
 RL Microbiology 143:2939-2943(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Mozer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetel D., Porwollik S., Prescott A.M.,

RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Terpatra P., Tognoni A.,
 RA Tsato V., Uchiyama S., Vanderbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler G., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Darchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis".
 RL Nature 390:249-256(1997).
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)
 or send an email to license@isb-sib.ch).
 CC
 DR ENBL; Y11043; CAA71937.1; -;
 DR ENBL; U93876; AAB80898.1; -;
 DR ENBL; Z93117; CAB14615.1; -;
 DR PIR; E69611; E69611.
 DR HGSP; Q03441; LOXA.
 DR Subtilisin; BGL1929; CYP4.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME P450; 1.
 DR Oxidoreductase; Monooxygenase; Heme; Complete proteome.
 KM METAL 359 359 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 SQ SEQUENCE 410 AA; 47384 MW; 035E9858A4761AB CRC64;
 Query Match 32.1%; Score 658; DB 1; Length 410;
 Best Local Similarity 37.4%; Pred. No. 3.9e-35;
 Matches 146; Conservative 78; Mismatches 136; Indels 30; Gaps 8;
 QY 1 MGVINPRQALQRLNGKQDAYHFFPWYESMRKDA---PVSFDEENQVWSVFLYDDWK 57
 Db 9 VTILTESQLSSRAF-----KDEAYE---FYKELRKSQALYPLSLGALGKGLISRYDDAI 60
 QY 58 KVVGDKEI-----PSS-----CMPQQTSSIGNSIINMDPPKHTKIRSVNKAFTPRVM 105
 Db 61 HLLKNEKLKNYENVETAKEKPAALLKNSEETLTKMNSDPPDNNRDLTLVQKAFTHMI 120
 QY 106 KQWEPRIQBITDELQKQFQSRSEFDLVHDSYPLPVIVISBELGVPSAHMQFKWASDL 165
 Db 121 LQLEDKIQHIADSLDKVQPNKFMNLVDVYAPFLPIIVISEXMLGIPLEDKQKFRVWSQAI 180
 QY 166 VSTPKDKSEAEKAFLEERDKCEELAAFFAGLIEKRNKPKQDIIISILVBAETGEKLS 225
 Db 181 I-----DFSDAPER--LQENDHLIGFVEYLSLVKRRREPAGDLISALIOASEGTQLS 234
 QY 226 GBELIPFCTLLVAGNETTTNLISNAMYSLBTPGVYEEELRSHPELMPQAVEALFRAP 285
 Db 235 TBEISYMLMLLVAGHETTVNLITNTYALMCHHQLKRLQCPDLMNSAIEALFHPSP 294
 QY 286 APVLR-RIAKRDTGIGGHLIKEGDMVLAFVANSRDEAKFDRPHMFDIRHPNPHIAFGH 344
 Db 295 VELTTIRWTAEFFILHQBQIEKRDVIVIIISANRDEKIFPNADIPDIERKKNRHIAFGH 354
 QY 345 GHFCLGAPLARLENIALTSLISAPPEME 374
 Db 355 GNHFCLGAPLARLEAKIAISTLLRRCPNIQ 384

RESULT 6
 CPYK SACER
 ID CPYK_SACER STANDARD; PRT; 405 AA.

or send an email to license@sib-sib.ch.

```

CC CC
CC EMBL; U51868; AAB17462.1; --
CC EMBL; A2008226; AAC00266.1; --
CC EMBL; Z99119; CAB14997.1; --
CC DR PIR; G69594; G69594.
CC DR HSP; Q55080; 1107.
CC DR Subtilist; BG11528; biof.
CC DR InterPro; IPR001128; Cytochrome_P450.
CC DR Pfam; PF00067; P450; 1.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00866; CYTOCHROME_P450; 1.
CC KW Biotin biosynthesis; Oxidoreductase; Monooxygenase;
CC Electron transport; Heme; Complete proteome.
CC METAL 345 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SEQUENCE 395 AA; 44865 MW; E4AC3AF2637ACE1A CRC64;
CC
CC Query Match 29.2%; Score 600; DB 1; Length 395;
CC Best Local Similarity 35.5%; Pred. No. 1.9e-31;
CC Matches 128; Conservative 84; Mismatches 131; Indels 18; Gaps 7;
CC
CC QY 25 HPFPYENRMDAPV---SFDEENQVWSVELYDDVKVGVGDKEL-----FSSCMPOQT 74
CC DB 15 NPSFYDTLRVHPIYKGSF-LKPGVWVGTGEETAALKDARFKVUTLPESSTKYQDL 73
CC
CC QY 75 SSGNS-IINMDPPKTKIRSVNKAFTPRVMKQWEPRIQETIDELIQFGSRSEFDLVH 133
CC DB 74 SHVQNMWLFQNPQDHRRLRTLASGAFRTTESYQPIIETVHLLDQVQKKRMEVIS 133
CC
CC QY 134 DFSYPLPVIVISLGVSAHMEQKANSLLVST-PRDKSEAKAFLEERDKCEELA 192
CC DB 134 DEAFPLASVIANIIGVPEDEQKWAASLIQIDFTSRKA----LTEGNMVAQAM 189
CC
CC QY 193 AFEAGLIEKRNKPDQIISILVEAEETGKLSGELIPFCITLLVAGNETTINLSNAM 252
CC DB 190 AFFKELIQKRNKQPDQWISMLKREK-DKLTREAASTCILLAGHETTVNLSNV 248
CC
CC QY 253 YSILETPGVYELRSHPELMPQAVEALFRAPAVLRRIAKRDTTEIGGHLIKEGDMVLA 312
CC DB 249 LCLLQHPQLLKRENPDLIGTAVECLRYESPTQMTARVASEDIDCGVTIRQGEQVYL 308
CC
CC QY 313 FVASNRDEAKFDRHMDIRHNPHTAFGHIHFCGAPLARLEANNIALTSLISAFPH 372
CC DB 309 LLGAANDPESITFNPEDFIDTRSPNPHLSFGHGHVCLGSSSLARLEAQAINTLQRMPS 368
CC
CC QY 373 M 373
CC DB 369 L 369
CC
CC RESULT 8
CC CPXJ_SACER STANDARD; PRT; 397 AA.
CC AC P48635; O33990;
CC DT 01-FEB-1996 (Rel. 33, Created)
CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Cytochrome P450 113A1 (Erythromycin B/D C-12 hydroxylase).
CC GN ERYK OR CYP113A1
CC OS Saccharopolyspora erythraea (Streptomyces erythraeus).
CC OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC OC Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
CC OX NCBI_TaxID=1836;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-NRRL 2338;
CC RX MEDLINE=97390129; PubMed=9249068;
CC RA Pereda A., Summers R., Katz L.;
CC RA "Nucleotide sequence of the ermE distal flank of the erythromycin
CC biosynthesis cluster in Saccharopolyspora erythraea.";
CC RT Gene 193:65-71(1997).
CC CC
CC -!- FUNCTION: Responsible for the C-12 hydroxylation of the
CC macrolactone ring of erythromycin.
CC -!- PATHWAY: Erythromycin B and D biosynthesis; final hydroxylation
CC step.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U82823; AAC45584.1; --
CC DR PIR; B40634; B40634.
CC DR HSP; Q55080; 1107.
CC DR InterPro; IPR001128; Cytochrome_P450.
CC DR Pfam; PF00067; P450; 1.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00866; CYTOCHROME_P450; 1.
CC KW Oxidoreductase; Monooxygenase; Electron transport; Heme;
CC Antibiotic biosynthesis.
CC FT METAL 339 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SQ SEQUENCE 397 AA; 43759 MW; 8306AF79414C981E CRC64;
CC
CC Query Match 27.8%; Score 571; DB 1; Length 397;
CC Best Local Similarity 39.2%; Pred. No. 1.4e-29;
CC Matches 136; Conservative 49; Mismatches 152; Indels 10; Gaps 5;
CC
CC QY 29 WTESNRKDPVSEFDEENQVWSVELYDDVKVGVGDKELFSS---CMPQQTSSGNSIINMD 85
CC DB 19 WLGTRKQPV-WQDRYGVWVFRHADVTQVLRDTATFSSDPTRVIEGASPTPGMIEHD 77
CC
CC QY 86 PKKHTKIRSVNKAFTPRVMKQWEPRIQETIDELIQFGSRSEFDLVHDFSYPPLVIVIS 145
CC DB 78 PEHRAIRKVSASATPTRTISDLPRIDVTRSLA--DAGESFDLVDLVLAFLPPTIVA 135
CC
CC QY 146 ELGVPSAHMEQKANSLLVSTPKDSEAKAFLEERDKCEELAAPFAGIIEKRNK 235
CC DB 136 ELGLPXPXDEHQFGWVGALVDIQMD--DPTPALAERIAVLNPLTAYLKARCAERRAD 133
CC
CC QY 206 PQDIIISILVEAEETGKLSGELIPFCITLLVAGNETTINLSNAMYSILETPGVYVEL 265
CC DB 194 PGDDLISRLVLAEDVGRALDDEEANEFTALLAGHITTVLGNIVRTLDEHPAHWDAA 253
CC
CC QY 266 RSHPELMPQAVEALFRAPAVLRRIAKRDTTEIGGHLIKEGDMVLAFLVASNRDEAKFD 325
CC DB 254 AEDPGRIPAIVEELVRYRPPQMQRTTKATEVAGVPIPAQVMVNTVLSANRSDAHD 313
CC
CC QY 326 RHEMFDIRHNP--PHIAPGHGHIHFCGAPLARLEANNIALTSLISAF 370
CC DB 314 DPDRFDPKSGGAAQSFSGHGVHFCGAPLARLENRVALEEIIARF 360
CC
CC RESULT 9
CC CPXJ_SACER STANDARD; PRT; 404 AA.
CC ID CPXJ_SACER
CC DT 01-FEB-1994 (Rel. 28, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE 6-deoxyerythronolide B hydroxylase (EC 1.-.-.-) (6-DEB hydroxylase)
CC DE (Erythromycin A biosynthesis hydrolase) (Cytochrome P450 107A1)
CC DE (CYP411A1) (P450eryf).
CC GN ERYF OR CYP107A1.
CC OS Saccharopolyspora erythraea (Streptomyces erythraeus).

```

IRON (HEME AXIAL LIGAND)		Q -> QDQ (IN REF. 1).		G -> R (IN REF. 1).	
FT	METAL	351	351		
FT	CONFLICT	263	263		
FT	CONFLICT	302	302		
FT	STRAND	6	6		
FT	TURN	7	8		
FT	HELEX	10	12		
FT	TURN	13	13		
FT	HELEX	16	26		
FT	STRAND	29	34		
FT	TURN	35	36		
FT	STRAND	37	42		
FT	HELEX	45	53		
FT	TURN	55	56		
FT	STRAND	57	58		
FT	TURN	61	62		
FT	TURN	66	67		
FT	HELEX	73	75		
FT	TURN	76	77		
FT	HELEX	80	86		
FT	HELEX	90	92		
FT	TURN	95	96		
FT	HELEX	97	106		
FT	TURN	107	108		
FT	HELEX	111	115		
FT	TURN	116	117		
FT	HELEX	118	132		
FT	STRAND	137	139		
FT	HELEX	140	143		
FT	TURN	144	146		
FT	HELEX	147	157		
FT	TURN	158	158		
FT	TURN	161	165		
FT	HELEX	166	175		
FT	HELEX	179	181		
FT	HELEX	182	205		
FT	HELEX	211	217		
FT	STRAND	219	220		
FT	TURN	221	223		
FT	STRAND	224	225		
FT	HELEX	228	259		
FT	HELEX	261	269		
FT	HELEX	271	273		
FT	HELEX	274	284		
FT	STRAND	291	295		
FT	STRAND	299	301		
FT	TURN	302	303		
FT	STRAND	304	306		
FT	TURN	308	309		
FT	STRAND	311	314		
FT	HELEX	316	319		
FT	TURN	320	321		
FT	TURN	323	325		
FT	TURN	329	330		
FT	TURN	334	335		
FT	TURN	339	340		
FT	TURN	343	344		
FT	TURN	348	349		
FT	TURN	352	353		
FT	HELEX	354	371		
FT	TURN	373	374		
FT	STRAND	376	377		
FT	HELEX	381	383		
FT	STRAND	386	386		
FT	STRAND	395	395		
FT	STRAND	399	401		
FT	SEQUENCE	404 AA;	45099 MW;		
FT	SEQUENCE	257ABEFC2D8B3FE	CRC64;		

```

Query Match      27.1%; Score 557; DB 1; Length 404;
Best Local Similarity 33.9%; Pred.No. 1.le-28;
Matches 130; Conservative 71; Mismatches 134; Indels 48; Gaps 11;

QY    20 KQDAYHFFWYS---MRKDPVSY--FQEENQVWSVFLYDDVKVVGDKELFSSCMQQ 73
       : :: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 8 ESDSPH-VDMYNTYAEALRETAPVTVRFUGQD-AWLVTGYDEAKAALSLRLSSD--PKK 63
QY 74 -----TSSIGNSIINMDPPKHKTIRSVVNVKAFTRPVAKQWEP 110
Db 64 KYPGVEVEFPAYLGPEDVNVYFATNMGTG-----DPPTHRLKLVQSEFTVRRVEAMP 119
QY 111 RIQETIDELIQKQGRSEFDLVHDSYPLVPVIVISELLGVPSAHMEQFVAM-SDLLVSTP 169
Db 120 RVEQITABELLDEWGSVVDIVDRFAHPLPIKVICELLDGDKYRGEGFGRWSSEILVMDP 179
QY 170 KDKSEAKAFLEERDKCEEELAAFPAGIIEKRNKPEODIISILVEAETGE-KLSGEE 228
Db 180 -----EKA--EQRGQAAREVNFILDIVERKRETEPGDLSALIRVQDDDDGRLSADE 230
QY 229 LIPFCTLLIVAGNETTNLISNAMYSLILETPGYEELRSHPELMPQAVREALFRFAPAV 288
Db 231 LTSIALVLLAGFEASVSLIGITGYLLTHPQALVRRDPSALPNAVEILRYAPPET 290
QY 289 LRIAKRDTIEGHLIKEDGMVLAFAVANRDEAKFDRPHMDIRHPNPHAFAGHGHF 348
Db 291 TTRFAAEVEIGGVAIPQYSTVLVANGAANRDPKQFPDRFDVTRDTRGHLISFGQGIHF 350
QY 349 CUGAPLARLEANIATLSLISAPP 371
Db 351 CMGRPLAKLEGEVALRGLFGRPP 373

RESULT 10
CPXW SULSO
ID CPXW SULSO STANDARD; PRT; 368 AA.
AC Q5080;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 119 (EC 1.14.14.-).
CN CYP119.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=ATCC 35091 / DSM 1616 / P1;
RC MEDLINE=96197795; PubMed=8617361;
RA Wright R.L., Harris K., Solow B., White R.H., Kennedy R.H.;
RT "Cloning of a potential cytochrome P450 from the archaeon Sulfolobus solfataricus";
RL FEBS Lett. 384:235-239 (1996).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99032823; PubMed=9813164;
RA McLean M.A., Maves S.A., Weiss K.E., Kreplich S., Sligar S.G.;
RT "Characterization of a cytochrome P450 from the acidothermophilic archaea Sulfolobus solfataricus";
RL Biochem. Biophys. Res. Commun. 252:166-172 (1998).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20261529; PubMed=10799487;
RA Koo L.S., Tschirret-Guth R.A., Straub W.E., Moenne-Loccoz P.,
RA Loehr T.M., Ortiz de Montellano P.R.;
RT "The active site of the thermophilic CYP119 from Sulfolobus solfataricus";
RL J. Biol. Chem. 275:14112-14123 (2000).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=20414776; PubMed=10957637;
RA Park S.-Y., Yamane K., Adachi S.-I., Shiro Y., Weiss K.E.,
RA Sligar S.G.;
RT "Crystallization and preliminary X-ray diffraction analysis of a cytochrome P450 (CYP119) from Sulfolobus solfataricus";
RL Acta Crystallogr. D 56:1173-1175 (2000).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.93 ANGSTROMS).

RX MEDLINE=20469464; PubMed=10859321;
RA Yano J.K., Koo L.S., Schuller D.J., Li H., Ortiz de Montellano P.R.,
RA Poulos T.L.;
RT "Crystal structure of a thermophilic cytochrome P450 from the archaeon Sulfolobus solfataricus";
RL J. Biol. Chem. 275:31086-31092 (2000).
CC -!- FUNCTION: CATALYZES THE H(2)O(2)-DEPENDENT EPOKIDATION OF STYRENE, CIS-BETA-METHYLSYRENE, AND CIS-STILBENE WITH RETENTION OF STEREOCHEMISTRY.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- MISCELLANEOUS: Stable at up to 85 degrees Celsius.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -!- CAUTION: DOES NOT SEEM TO EXIST IN THE COMPLETE GENOME OF P2 STRAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; U51337; AAS03278.1; -;
DR PIR; S71328; S71328.
DR PDB; 1F4T; 23-OCT-00.
DR PDB; 1F4U; 23-OCT-00.
DR PDB; 1IO7; 21-JAN-03.
DR PDB; 1IO8; 21-JAN-03.
DR PDB; 1IO9; 21-JAN-03.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Iron; Heme;
KW 3D-structure.
FT ACT_SITE 213 213 IRON (HEME AXIAL LIGAND).
FT METAL 317 317
FT HELIX 2 11
FT STRAND 14 16
FT STRAND 21 23
FT HELIX 25 34
FT TURN 36 38
FT STRAND 39 40
FT HELIX 46 53
FT TURN 54 56
FT HELIX 63 65
FT HELIX 68 70
FT TURN 73 74
FT HELIX 75 80
FT HELIX 81 83
FT TURN 84 87
FT HELIX 89 108
FT TURN 109 109
FT TURN 112 113
FT STRAND 115 116
FT HELIX 117 123
FT TURN 121 123
FT HELIX 124 134
FT TURN 135 135
FT HELIX 138 140
FT HELIX 141 147
FT HELIX 148 150
FT TURN 152 153
FT TURN 158 159
FT HELIX 161 178
FT TURN 179 180
FT HELIX 184 190
FT TURN 191 191
FT HELIX 196 208
FT TURN 209 210
FT HELIX 211 227
FT TURN 228 229

```

FT 231 237
FT 238 238
FT 240 250
FT 254 254
FT 254 254
FT 257 261
FT 265 267
FT 268 269
FT 270 272
FT 274 275
FT 277 280
FT 282 285
FT 286 287
FT 289 291
FT 293 296
FT 300 301
FT 309 310
FT 313 315
FT 318 319
FT 320 335
FT 336 337
FT 340 348
FT 354 354
FT 357 365
SQ SEQUENCE 368 AA; 42863 MW; E2A5C4F064537E8C CRC64;

Query Match      26.7%; Score 548.5; DB 1; Length 368;
Best Local Similarity 34.7%; Pred. No. 3.4e-28;
Matches 134; Conservative 68; Mismatches 145; Indels 39; Gaps 10;

QY 27 PPWYESMRKAPVGFDEENQVWSVFLYDDVKKVGVGDKELFSKMP---QQSSSTGN----- 79
Db 2 YDWFSEMRKKDPVYD--GNWQVFSYRYTKVELNLFNFKSFDLTGYHERLEDLNGKIR 59
QY 80 -----SIINMDPPKTKIRSVNKAFTPRVMKOWEPRIQETIDELIQKQSERFLV 132
Db 60 FDIPTRTVMTSDPLDDELMSGADIFSPQKQTLFTFIRETRSLDSDIDPR-EDDIV 118
QY 133 HDFSYPFLVIVISLLGVPSAHMEQFKAWSLLVSTPKDKSE--EAEKAFLEKRDCKEE 190
Db 119 KKLAVPPIIIVISKILGLPTBDEKFKEMSDLVAFRLGKPGEIFELGKKYLE----- 170
QY 191 LAFFAGIIEBKRNKPDQDIISILVEAETGEKLSGBELPFCILLVAGNETTNLISN 250
Db 171 -----LIGYVKDHLNS-GTEVVSRYVNS-----NLSDIEKLGYYIILLTAGNETTNLISN 220
QY 251 AMYSILTEPGVSELRGHPMLPOAVEALRFAPAPVLRRIAKRDTBIGGHLIKESDMV 310
Db 221 SVIDFTRF-NWQRIREE-NLYLKAIEALAYSPPVMTVTKERVKLGQTIIESEYV 278
QY 311 LAFVANSRDEAKFDRPHMFDIRHPNPHIAFGHIFCLGAPLARLEANTALSLISAF 370
Db 279 RVWIASANRDEEVPHDGEKTEPDNRPNPHLSFGSGIHLCLGAPLARLEARIAIEBFKRF 338
QY 371 PHMECVSITPIENSVIYGLKSFYKVM 396
Db 339 RHEILDTEKVPNEVLNGYKRLVRL 364

RESULT 11
C124_MYCTU
ID C124_MYCTU STANDARD; PRT; 428 AA.
AC Q50696;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative cytochrome P450 124 (EC 1.1.14.-.-).
GN CYP124 OR RV2266 OR MT2328 OR MTCY339.44C OR MB2289.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN 1;

```

```

RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva Y., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parrhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; Z77153; CAB00969.1; -.
DR EMBL; AE007076; AAK46610.1; -.
DR EMBL; BX248342; CAD97150.1; -.
DR PIR; F70729; F70729.
DR HSSP; P33006; 1CPT.
DR TIGR; MT2328; -.
DR Tuberculist; Rv2266; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME P450; FALSE NEG.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 379 379 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 23 23 D -> G (IN REF. 2).
FT CONFLICT 75 75 Y -> N (IN REF. 2).
SQ SEQUENCE 428 AA; 47824 MW; 76B1F3CSAE348591 CRC64;

Query Match      25.3%; Score 520; DB 1; Length 428;
Best Local Similarity 32.1%; Pred. No. 2.7e-26;
Matches 126; Conservative 82; Mismatches 148; Indels 36; Gaps 10;

QY 30 YESMRKAPVS-----EDEENQVMSVFLYDDVKKVGVGDKELFS-----CMFQQTSS 76
Db 43 FATLRREAPISFWFTIELPGFVAGHWAHTKYDDVFYASRHPDIFSSYPNTINDQIPE 102
QY 77 IG---NSTINMDPPKTKIRSVNKAFTPRVMKOWEPRIQETIDELIQKQGRS---EFD 130
RN 1;

```

Db 103 LAEYFGSMVLDPRHQRLRSIVSRAFTPKVAVIAEAAVRAHRLVSSMIANRPDQAD 162
Qy 131 LVHDFSPLPVIVISSELLGVSAHMEQKAWSDLLVS-TPDKSREAKAFLERDKCEE 189
Db 163 LVSELAGPLPQIICDMGIPKADHQRIHFHWNIVLFGDPDLAVD-----FDFMQVSA 217
Qy 190 ELAAFFAGIIEKENKPEQDIISILVEAETGEKLSGELIPFCTLLVAGNETTNLIS 249
Db 218 DIGAYATAEDRRVNHDDLTSSIVAEVDERLSREIASPFLVAVAGNETTRNAT 277
Qy 250 NAMYSIETGCVYBELRSH-----PELMPQAVEARLPRAPVLRRIAKDTRIGGHLI 304
Db 278 HGVLALSR-----YPEQRDRWSDFDGLAPTAVEIVRWASFWYMRRTLTQDIELRGTKM 333
Qy 305 KEGDMVLAVASANDRAKDPDRPHMFDLRPHNPHIAP-GHGIFHCGLAPLARLEANAL 363
Db 334 AAGDKVLSWYCSARDESKFADPWTFDLARNPHLFGGGAHFCGLANLARIRVAF 393
Qy 364 TSLISAPFMBCV-SITPIENSVIYGLKSRV 394
Db 394 DELRRQMDVVATEEPARLLSQFINGIKTLFV 425
RESULT 12
C123 MYCTU STANDARD; PRT; 402 AA.
AC P77923;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative cytochrome P450 123 (EC 1.14.-.-).
GN CYP123 OR RV0766C OR MT0790 OR MTCY363.11C OR MB0789C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=92295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218336;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin S., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."

RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z80226; CAB02396.1; -.
DR EMBL; AE006970; AAK45032.1; -.
DR EMBL; BX248336; CAD93651.1; -.
DR PIR; A70707; A70707.
DR HSSP; Q00441; 10XA.
DR TIGR; MT0790; -.
DR Tuberculin; RV0766C; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 350 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 402 AA; 45421 MW; 76AD8430.9361798 CRC64;
Query Match 25.1%; Score 516; DB 1; Length 402;
Best Local Similarity 33.5%; Pred No. 4,6e-26;
Matches 120; Conservative 72; Mismatches 152; Indels 14; Gaps 4;
Qy 26 PPMYESMRKDAVPSFDEENQVSVLYDYDKVGVGDKELFS-----SCMPQOTSIG-- 78
Db 23 PYPYVRLRDEAPLYRNEERNFVAVSRHVDLQGPDRDSTALSNAYGVSLDPSRSTSEAYR 82
Qy 79 -NSIINMDPPKTKLRVYNKAFTRVNMKQEPRIQETIDELIQFGQSESEDLVDHSY 137
Db 83 VMSMLAMDPPAHLRMTLVSKGFTPRIRIE-EPQVLELARLHLSALQTESDFVAEFAG 142
Qy 138 PLPVIVISELLGVSAHMEQKAWSDLLVSTPKDSEAEAKAFLERDKCEEELAAFFAG 197
Db 143 KLPMDVISELIGVPTDTRARITALADAVLHREDGVADVPFPAY-----AASIELMRYVAD 197
Qy 198 IIEEKNRPEQIISILVEAETGEKLSGELIPFCTLLVAGNETTNLISNAMYSLILE 257
Db 198 LIAEFRFRPANNLTALLAAELDGRLSQDQIMAFLEFLMVIAGNETTTKLANAVYWAH 257
Qy 258 TPGVYEELSRHPQLMPQAVEARLPRAPVLRRIAKDTRIGGHLIKEGDMVLAFVSA 317
Db 258 HPGQLARVFAHRSRIPMWVEETLYDTSSQILARTVAHDLTYDTTPEGEVILLPGSA 317
Qy 318 NRDEAKFRPHMFDIRRHEN-PIAIFGHIHFCGLAPLARLEANALTSLSIAFFPME 374
Db 318 NEDRVFDDPDYRIGREIGCKLVSGSGAHCGLANLARIRVAF 375
RESULT 13
C142 MYCTU STANDARD; PRT; 398 AA.
AC O53563;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative cytochrome P450 142 (EC 1.14.-.-).
GN CYP142 OR RV3518C OR MT3619 OR MTV023.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RA Nature 393:537-544(1998).
RA [2]
RA SEQUENCE FROM N.A.
RA STRAIN=CDC 1551 / Oshkosh;
RA MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RA J. Bacteriol. 184:5479-5490(2002).
RA [1]
RA -!- SIMILARITY: Belongs to the cytochrome P450 family.
RA This SWISS-PROT entry is copyright. It is produced through a collaboration
RA between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA the European Bioinformatics Institute. There are no restrictions on its
RA use by non-profit institutions as long as its content is in no way
RA modified and this statement is not removed. Usage by and for commercial
RA entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RA or send an email to license@isb-sib.ch).
RA [1]
RA EMBL; AL022022; CAAL7755.1; -;
RA EMBL; AE007165; AAK47979.1; ALT_INIT.
RA PIR; H70807; H70807.
RA HSSP; P33006; LCPT.
RA TIGR; MT3619; -;
RA Tuberculist; RV3518c; -;
RA InterPro; IPR001128; Cytochrome_P450.
RA Pfam; PF00067; P450; 1.
RA PRINTS; PR00385; P450.
RA PROSITE; PS00086; CYTOCHROME_P450; 1.
RA Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
RA Complete proteome. 340 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
RA METAL 340 MTEAPVDVLAQGNFYASR -> GWSFNSRAPPSPARRP
RA CONFLICT 1 18 {IN REF. 2}.
RA SEQUENCE 398 AA; 44398 MW; BCF3C23ECB5767F CRC64;
RA Query Match 25.1%; Score 514.5; DB 1; Length 398;
RA Best Local Similarity 34.2%; Pred. No. 5.6e-26;
RA Matches 129; Conservative 61; Mismatches 164; Indels 23; Gaps 7;
QY 30 YESRKDAPVDFERNQVMSVFLYDDVKKVGVGKELFSSC---MPQQTSSIGNSIINMDP 86
DB 24 YRWKRNQPV FRDNGLAASSTYQAVIDAEPQLFNSNAGGIRPDQPAL--PMMLMD 80
QY 87 PKHTKIRSVYKAFTRPVKWKQWEPRIQETDLELOKFGQRSGFDLVHDFSYELPVIVISE 146
DB 81 PHLRLRRKLNVNAGTRKRVKDKKASIAALCDTLIDAVCGERGCDVFDRLAAPLPMVAVIGD 140
QY 147 LIGVPSAHMEQFANSLLVLS-----TPKDKSEAEKAFLEERDKCEBELAFAFFACIES 201
DB 141 MLGVPRPEQDMFLRWSDDLWTLFSSVSGEDFQITWDAFAAYND-----FTRATIAA 192
QY 202 KENKPERQDIISILVRAEETGEKLSGELIPFCTILLVAGNETTNNLISNAMYSILETPGV 261
DB 193 READPTDULVSVLVSSEVNGERLSDELVWETLLIIGGDETRHTLSGTEQLLRNQ 252
QY 262 YEELRSHPELMPQAVEALFRAPAPVLRARIKRDTEIGHLIKEDGMVLAFFVANSRDE 321
DB 253 WDLQDPDLSLLGALIEMLRWTAFAVNMCKRVLTADTEFFGTALCAGEKMWLLPESANFDE 312

QY 322 AKFDRPHMDIRRHNPHTAFGHGHCFCGAPLARLEANIALTSLISAFPHMECV---SI 378
DB 313 AVFCEPEKEDVQVQNPNSHLAPFGTHFCGNGQLARLELSMTXERVLRPLDLRLVADDSV 372
QY 379 TPTE-NSVIYGLKSRV 394
DB 373 LPLRPANFVSGLESMFV 383
RESULT 14
C125_MYCTU STANDARD; PRT; 433 AA.
ID C125_MYCTU
AC P71856;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative cytochrome P450 125 (EC 1.14.-.-).
GN CYP125 OR RV3545C OR MT3649 OR MTCY03C7.11 OR MB3575C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RN SEQUENCE FROM N.A.
RN SPECIES=M.tuberculosis; STRAIN=H37Rv;
RC MEDLINE=9634230;
RC Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RC Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RC Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RC Davies R., Devlin K., Krogh A., McLean J., Moule S., Murphy L.,
RC Hornsby T., Jagels K., Krogh A., McLean J., Rajandream M.A., Rogers J.,
RC Oliver S., Osborne J., Quail M.A., Rajandream M.A., Squares R.,
RC Rutter S., Seeger K., Skelton S., Squares S., Barrall B.G.;
RC Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RC "Deciphering the biology of Mycobacterium tuberculosis from the
RC complete genome sequence.";
RC Nature 393:537-544(1998).
RN [2]
RN SEQUENCE FROM N.A.
RN SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RC MEDLINE=22206494; PubMed=12218036;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RC Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RC Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RC Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RC Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RC "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RC laboratory strains.";
RC J. Bacteriol. 184:5479-5490(2002).
RN [3]
RN SEQUENCE FROM N.A.
RN SPECIES=M.bovis; STRAIN=AF2122/97;
RC MEDLINE=22709107; PubMed=12788972;
RC Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RC Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RC Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RC Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RC "The complete genome sequence of Mycobacterium bovis.";
RC Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC [1]
CC EMBL; Z82098; CAB05061.1; -;
CC EMBL; AE007166; AAK48008.1; -;

[illegible]

Wed Jun 2 13:01:00 2004

Search completed: May 28, 2004, 23:18:05
Job time : 42 secs

us-09-869-334b-1.rsp

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 18:12:54 ; Search time 3275 Seconds
(without alignments)
15762.366 Million cell updates/sec

Title: US-09-869-334B-2

Perfect score: 1191

Sequence: 1 atgaatgtgttaaacgcgcg.....gcttcggtgaaatgtaa 1191

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pi:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1189.4	99.9	35739	1	AF015825 Bacillus
2	1189.4	99.9	201241	1	BSUB00007 Z99110 Bacillus su
3	425	35.7	966	6	AX433775 Sequence
4	242	20.3	1213	6	AX433767 Sequence
5	234.4	19.7	290685	1	AE017032 Bacillus
6	229.8	19.3	3187	1	M24523 B.subtilis
7	182.6	15.3	4317	1	BACRIP Z21972 B.megateriu
8	169	14.2	1902	1	BMCP4501 X11610 Bacillus me
9	165.8	13.9	6629	1	AB072568 Streptomy
10	162.4	13.6	5608	1	BSCYPAZL Y11043 Bacillus su
11	162.4	13.6	17828	1	BSU93876 U93876 Bacillus su
12	162.4	13.6	199173	1	BSUB0014 Z99117 Bacillus su
13	159.6	13.4	300893	1	AE017006 AE017006 Bacillus
14	156.4	13.1	10167	1	SFU08223 U08223 Streptomyce
15	150	12.6	342300	1	MLEPRTN8 A1583924 Mycobacte
16	149	12.5	1233	6	AX697983 Sequence
17	149	12.5	60196	6	AX697977 Sequence
18	145.8	12.2	17512	1	AB071405 Lechevali
19	145.8	12.2	25681	1	SAE414559 AJ414559 Saccharot
20	145.8	12.2	26144	1	AB090952 Lechevall
21	145.8	12.2	28654	1	AF534707 Lechevall
22	144.2	12.1	3078	1	SRCPC450A M83110 Saccharopol
23	144	12.1	39228	1	MLCB1788 AL008609 Mycobacte
24	139.2	11.7	8478	6	AR173058 Sequence
25	139.2	11.7	10153	1	BSU51868 US1868 Bacillus su
26	139.2	11.7	199922	1	BSUB0016 Z99119 Bacillus su
27	139.2	11.7	220060	1	AF008220 AF008220 Bacillus
28	138.2	11.6	207829	1	BSUB0010 Z99113 Bacillus su
29	136.4	11.5	6739	1	AB088066 Bacillus
30	135.4	11.4	293175	1	AP005023 Streptomy
31	125	10.5	300350	1	AF006574 Gloeobact
32	124.2	10.4	145911	1	AF003014 Mesorhizo
33	122.4	10.3	309050	1	SCQ939117 AE011961 Xanthomon
34	120.6	10.1	10417	1	AE011961 AE011961 Xyella f
35	120.6	10.1	300885	1	AE012559 Agrobacte
36	119.2	10.0	10029	1	AE009087 Agrobacte
37	119.2	10.0	10184	1	AE008053 Agrobacte
38	118.4	9.9	154620	2	BX323047 Danio rer
39	118.2	9.9	16057	1	AE007165 Mycobacte
40	118.2	9.9	349606	15	BX842583 Mycobacte
41	117.8	9.9	24268	1	AB070947 Streptomy
42	117.8	9.9	308550	1	AP005030 Streptomy
43	117	9.8	8407	1	SE082823 U82823 Saccharopol
44	116.4	9.8	18109	1	AE006971 Mycobacte
45	116.4	9.8	320050	1	BX248336 Mycobacte

ALIGNMENTS

RESULT 1
AF015825
LOCUS AF015825 35739 bp DNA linear BCT 07-MAY-1998
DEFINITION Bacillus subtilis 168 cot1-rapA region sequence.
ACCESSION AF015825
VERSION AF015825.1 GI:2612880
KEYWORDS
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 35739)
AUTHORS Rivolta,C., Soldo,B., Lazarevic,V., Joris,B., Maue,C. and Karamata,D.
TITLE A 35.7 kb DNA fragment from the Bacillus subtilis chromosome containing a putative 12.3 kb operon involved in hexuronate

catabolism and a perfectly symmetrical hypothetical
catabolite-responsive element

JOURNAL
MEDLINE

98240225

9573062

2 (bases 1 to 35739)

REFERENCE

AUTHORS

Rivolta, C., Soldo, B., Lazarevic, V., Joris, B., Maue, C. and

Karamata, D.

Direct Submission

Submitted (24-JUL-1997) Institut de Genetique et de Biologie

Microbiennes, Rue Cesar-Roux 19, Lausanne CH-1005, Switzerland

Location/Qualifiers

1..35739

/organism="Bacillus subtilis"

/mol_type="genomic DNA"

/strain="168"

/db_xref="taxon:1423"

/map="cot-rapA region"

complement (<1..129)

/gene="cotJ"

complement (<1..114)

/gene="cotT"

/codon_start=1

/transl_table=11

/product="spore coat protein"

/protein_id="AAC46305.1"

/db_xref="GI:2612881"

/translation="MDYPLNEQSFEQITPYDERQPYYPYPRPPFPYYP"

complement (122..129)

/gene="cotT"

355..1806

/gene="Yjfa"

355..361

/gene="Yjfa"

368..1771

/gene="Yjfa"

/note="similar to Rhizobium leguminosarum NodB protein"

/codon_start=1

/transl_table=11

/product="NodB-like protein"

/protein_id="AAC46306.1"

/db_xref="GI:2612882"

/translation="MLAKRIKPHVLIADVGVVGLIGFHHSLKKEKTVNKKVTRDSQ

YGNVETAVNCKTQVAVYVFNKEMDSALKRFAKVRQFQKQKTVQDQHTT

KNELNDVYKIVHAKQTVAVVFNKEMDSALKRFAKVRQFQKQKTVQDQHTT

EDADYHLKSLIAYHSLKKNKIDAADALLKEGTAKKNSFRAIKEDYIELVFTY

QAAVYLGSGSTAKKSLKIDKQYIDKAKNKIKKQKPKHEVISLPERIVDPN

QKVIATFDQGNPATNQIDLSKKYKGHATFFVIGSRVQYYPETILRMKEGNEVG

NHWSHPLTLRLSVKALQINDTODIIEKISGRPTLVPPYGGINDELRSQMDV

ALWDVDPEDKQNKKTIVDRVNMVQAGDGRITLIDHIRTSAADADEIINKLTDQGYQ

LTVTSQLEEVKKQREAK"

1775..1806

/gene="Yjfa"

complement (1811..2298)

/gene="Yjfa"

complement (1811..2284)

/gene="Yjfa"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAC46307.1"

/db_xref="GI:2612883"

/translation="MKRLPKASLMLFAVVFVAVKAPAKAETHAYDGKSPYNDCA

SSGSTKSSNLVNASQVIGVSLKSSCTCKTAWAKITMNTLTSGVANAETRTND

GKKYNCDSAGNGKAVAGKSCCTPMWYDLDPRTSFAFKYSGPNLWATTGSY"

complement (2291..2298)

/gene="Yjfa"

complement (2363..2589)

/gene="Yjfa"

complement (2363..2399)

/gene="Yjfa"

complement (2409..2576)

/gene="Yjfa"

RBS
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC46308.1"
/db_xref="GI:2612884"
/translation="MDIPALSVMHQASLAQNVIATKKRLDTAQNADOTLKMTOH
PTLQGTIDVKA"
complement (2581..2589)
/gene="Yjfb"
2889..3632
/gene="Yjfc"
2889..2896
/gene="Yjfc"
2703..3632
/gene="Yjfc"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC46309.1"
/db_xref="GI:2612885"
/translation="NNLVMEKTFEQYEKLFSEEQKREDVRYTMMRPFKMTAIQV
PLKAPNGYDVIMAAKMLGYLDVDAESQKALQILKESHVSETAESALRQCIFAE
REOLRVNAKIKFGLVADPNKLQKQKXGFGGIPIGTHVWNPNSYMLPRIPSIIA
HEFHNVRSYIDFHGGSVSGDYLYIEGLAESFARELEGTRELGPWTFDEHLYQ
SIDVINEVLDRKFSVSYMFQDPIANDQGFAPGLSAPAGYAVGHAVQSFNQH
ITISEATRLDARTIISQGLFLHEKSVTVSGAF"
3602..3630
/gene="Yjfc"
complement (3610..4024)
/gene="Yjga"
complement (3610..4008)
/gene="Yjga"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC46310.1"
/db_xref="GI:2612886"
/translation="MKRNMIVAVFTILIGLGLGRASFVLPDTLNTVGLSMLAA
MFTGCGFLPRKLTMTIGTISLSCFCVFIHFSQLYEAEMIDQIDRTSLGLVLGYGFL
WSDIEATIGLACAALELLVLGKKRCM"
complement (4020..4024)
/gene="Yjga"
complement (4084..4697)
/gene="Yjgb"
complement (4084..4099)
/gene="Yjgb"
complement (4109..4684)
/gene="Yjgb"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="AAC46311.1"
/db_xref="GI:2612887"
/translation="MKMTSAITAAAVTSCFTGFGNAPSAPAKAACTNLSENTN
QSAELVKNLYNTAYKEMPPQAGGLTINKSTKGDVHAARFGEFVPGGNREDLHW
NMQPGYGSYHKDMTISEIRYFTGTVERQLNIGVTPVLOQLQGPVNVLTVPFTD
EIDYVDTGRYELHFVIGTIDQTAHVNLIKAK"
complement (4690..4697)
/gene="Yjgb"
4814..7787
/gene="Yjgc"
4814..4822
/gene="Yjgc"
4830..7787
/gene="Yjgc"
/note="similar to formate dehydrogenase alpha subunit of
Moorella thermoacetica; member of the 4Fe-4S ferredoxins
family"
/codon_start=1
/transl_table=11
/product="formate dehydrogenase alpha subunit homolog"
/protein_id="AAC46312.1"
/db_xref="GI:2612888"

/translation="MAGKTTIINGVEMERASQTVLQMLNSSIEVPQVCHPSLP
IFETCDIVSINGELXESCSARKLKGDDVITLSPDKKAOVIGMDKLLYNHLYCTVC
DYNGGCEIHNIVKEMKIHQSPIDPHKPKHDSHPFVYDPODCTICGRCEACQD
VQVETLTIDWEKRPVIMNDVPIINNESSCVSGHCSTVPCNMWMEKMEGEAGYL
TGNNETLRPMIEITKGVETGYSILAI SDMESAMRDERIKKTKTVCTYCGVGSFDV
WTKRDLKVEPOEAPANGISCTVKGKFGWDFNSEERLTKPLIREGDFHREABEB
ALLILAKFTLEKAPDPSLATITGSKTNEBSYLMQKLARGVIGTNNVDCSRVQC
SPATAGLFTVGVGSDSGS"TDIAQADLVLIIGSNTSESHPLVSTRKRAHKLGRQKV
IVADIIRHWAERSDILFVQPRAGSDIVWLNAIAKYLKENGKADERPLRVRNGRDEV
KSLAPYLEAERSTGIDQDSTLLIOMAEIMIGADSVCAALWAMGTUOHIGSDSTASLN
LLLVNIGYKPGAGSVPKPLHNSVQSGSPGMPDRUPGYEKVDEQVROKIERVNGV
PLPEKPMIWHMEIKIHSQKAMVYKGEMLVDSNIMHVAEYKELDFVYVQDIF
LSTAEADVVDLPASLEKGTPTNERRIQRYVFEPLGESKPDQWQIMEVANKL
GAGLVEHADIMEEAAKLSPIVAGTVYERLGVNSIQWPVWADKQDPLFTERFPF
PDGKALIPVQWTEPEKPEGYDLHVNGBLLSHFHEGNTLYKSGISEKTPVEFLBI
SPLEAERGIDQGLVRLTSPFGNKGKCLITRBYKKEVYLPMDNSGEAAILNLTGS
HADKDTDPAYKETSARKEILKHDGSPKPNHRNNGNPOQIGVQVHKWKARDYIF
PGDAVKRGMGNGH"
762. .8373
/gene="yjd"
762. .7769
/gene="yjd"
7780. .8340
/gene="yjd"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC46313.1"
/db_xref="GI:2612889"
/translation="MAKAIKRIKIEVTEEDQRKDRREIEDALIDHKEALETLMHL
GHMERGLVLLRGLFGQKVLIDLVKKADTEANTLKNLLLFGLGLMDVKOLE
PLILKVNAGVASAEVOKNSEKTYFOIIRSLKDPINKSITLFLSKGMGQDTKEL
ERTQPEHQKHQEPREKRGMKRD"
8347. .8373
/gene="yjd"
8525. .9223
/gene="yjb"
8525. .8530
/gene="yjb"
8537. .9178
/gene="yjb"
stem_loop
gene
RBS
CDS
Query Match 99.9%; Score 1189.4; DB 1; Length 35739;
Best local similarity 99.9%; Pred. No. 0;
Matches 1190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGAATGTGTTAAACCGCGGAGCGCTTGCAGGAGCGCTGCTCAATGGGAAACAAA 60
DB 10583 ATGAATGTGTTAAACCGCGGAGCGCTTGCAGGAGCGCTGCTCAATGGGAAACAAA 10642
QY 61 CAGGATGGTATCATCGGTTTCCATGTATGAATCGATGAGAGAGAGATGCGGCTGTTC 120
DB 10643 CAGGATGGTATCATCGGTTTCCATGTATGAATCGATGAGAGAGAGATGCGGCTGTTC 10702
QY 121 TTTGATGAAGAAACCAAGTGTGGAGCGTTTCTTTATGATGATGTCAAAAGCTGT 180
DB 10703 TTTGATGAAGAAACCAAGTGTGGAGCGTTTCTTTATGATGATGTCAAAAGCTGT 10762
QY 181 GGGGATAAAGAGTGTGTTTCCAGTTGATCGCGGAGCAGACAGCTCTATTGGAATTC 240
DB 10763 GGGGATAAAGAGTGTGTTTCCAGTTGATCGCGGAGCAGACAGCTCTATTGGAATTC 10822
QY 241 ATCATTTACATGGACCGCGGAGCAGATCAAAAATCCGTTTCACTGCTGTAACAAAGCTTT 300
DB 10823 ATCATTTACATGGACCGCGGAGCAGATCAAAAATCCGTTTCACTGCTGTAACAAAGCTTT 10882
QY 301 ACTCCGCGGTGATGAAGCAATGGGAACCGAGAAATTCAGAAATCAAGATCAAGATCAAT 360
DB 10883 ACTCCGCGGTGATGAAGCAATGGGAACCGAGAAATTCAGAAATCAAGATCAAGATCAAT 10942
QY 361 CAAAAATTCAGGGGCGAGTGAAGTTGACCTGTTTACGATTTTTCATACCGCTTCGG 420
DB 10943 CAAAAATTCAGGGGCGAGTGAAGTTGACCTGTTTACGATTTTTCATACCGCTTCGG 11002

QY 421 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGGCGAGATGGAACGCTTTTAAAGCA 480
DB 11003 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGGCGAGATGGAACGCTTTTAAAGCA 11062
QY 481 TGGTCTGATCTTCTGGTCACTACACCGAAGGATAAAAGTGAAGAGCTGAAAAGACCTTT 540
DB 11063 TGGTCTGATCTTCTGGTCACTACACCGAAGGATAAAAGTGAAGAGCTGAAAAGACCTTT 11122
QY 541 TTGGAAGAACGAGATAAAGTGTGAGGAAGAACTGSCGCGCTTTTTCGCGGATCATAGAA 600
DB 11123 TTGGAAGAACGAGATAAAGTGTGAGGAAGAACTGSCGCGCTTTTTCGCGGATCATAGAA 11182
QY 601 GAAAGCGAAACAAACCGGAACAGATATATTTCTATTTTAGTGAAGCGGAAGAAACA 660
DB 11183 GAAAGCGAAACAAACCGGAACAGATATATTTCTATTTTAGTGAAGCGGAAGAAACA 11242
QY 661 GGCGAGAGCTGTCGCGTGAAGAGCTGATTCGCTTTTGACGCTGCTGCTGCTGCTGCTG 720
DB 11243 GGCGAGAGCTGTCGCGTGAAGAGCTGATTCGCTTTTGACGCTGCTGCTGCTGCTGCTG 11302
QY 721 AATGAAACCACTACAAACCTGATTTCAAATCGGATGTACAGCATATTAAGAACGCGCAGC 780
DB 11303 AATGAAACCACTACAAACCTGATTTCAAATCGGATGTACAGCATATTAAGAACGCGCAGC 11362
QY 781 GTTTACGAGGAACCTGCGGAGCATCTGAACTGATGCTCAGGCACTGCGGAGGAGCGCTTG 840
DB 11363 GTTTACGAGGAACCTGCGGAGCATCTGAACTGATGCTCAGGCACTGCGGAGGAGCGCTTG 11422
QY 841 CGTTTCAGAGCGCGCGCGCGCTTTTGGAGCGCATGTCGCAAGCGGATACGAGATCGGAG 900
DB 11423 CGTTTCAGAGCGCGCGCGCGCTTTTGGAGCGCATGTCGCAAGCGGATACGAGATCGGAG 11482
QY 901 GGGCACCTGATTAAGAAAGGTGATATGTTTGGGCTTTTGGGCTTTTGGGCTTTTGGGCTTT 960
DB 11483 GGGCACCTGATTAAGAAAGGTGATATGTTTGGGCTTTTGGGCTTTTGGGCTTTTGGGCTTT 11542
QY 961 GAAGCAAGTTTGAAGACAGCGGATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 11543 GAAGCAAGTTTGAAGACAGCGGATGTTGATGATGATGATGATGATGATGATGATGATGAT 11602
QY 1021 GGGTTGGCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
DB 11603 GGGTTGGCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11662
QY 1081 ATGCGGTTAAAGCTGTTTGTGATTTCTGTTTCTGATGATGATGATGATGATGATGATGAT 1140
DB 11663 ATGCGGTTAAAGCTGTTTGTGATTTCTGTTTCTGATGATGATGATGATGATGATGATGAT 11722
QY 1141 ATTGAAACAGTGTGATATACGATTAAGAGCTTCCGTGTCGAAATGTAA 1191
DB 11723 ATTGAAACAGTGTGATATACGATTAAGAGCTTCCGTGTCGAAATGTAA 11773
RESULT 2
BSUB0007 201241 bp DNA linear BCT 07-JUL-2003
LOCUS Bacillus subtilis complete genome (section 7 of 21): from 1209742
DEFINITION to 1410982.
ACCESSION Z99110 AL009126
VERSION Z99110.2 GI:32468738
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Kunst, F., Ogasawara, N., Moser, I., Albertini, A.M., Allori, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borcher, S.,
Borriss, R., Boursier, L., Brans, A., Braun, M., Brigne, L.S.C.,
Bron, S., Brouillet, S., Bruch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Chci, S.K., Codani, J.J., Conner, I.F., Cummings, N.J.,
Daniel, R.A., Denicot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,

	Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A., Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J., Grandi,G., Guisepi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R., Henaud,A., Hilbert,H., Holtsappel,S., Hosono,S., Hullo,M.F., Ikeya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y., Klaert-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P., Koningsstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A., Lardiniois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H., Masuda,S., Mauei,C., Medigue,C., Medina,N., Mellado,R.P., Mizuno,M., Moesti,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M., Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,I.M., Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Fujic,P., Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M., Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T., Scanlan,E., Schleich,S., Schroeter,P., Scoffone,F., Sekiguchi,J., Sekowska,A., Seror,S.J., Serron,P., Shin,B.S., Soldo,B., Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K., Takeuchi,M., Tanakoshi,A., Tanaka,T., Terpetra,P., Tognoni,A., Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A., Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T., Winters,P., Wiptat,A., Yamamoto,H., Yamanaka,K., Yasumoto,K., Yata,K., Yoshida,K., Yoshikawa,H.F., Zumbstein,E., Yoshikawa,H. and Danchin,A.	The complete genome sequence of the gram-positive bacterium Bacillus subtilis	Nature 390 (6657), 249-256 (1997)	98044033 9384377	2 (bases 1 to 201241) Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A. Direct Submission Submitted (27-JUN-2003) I. Moezer, A. Danchin, Institut Pasteur, Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moezer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48 On Jul 7, 2003 this sequence version replaced gi:2633472. This entry contains data from release R16.1 of the Subtilist database. Further data on gene annotation and detailed information about changes from previous releases can be found at http://genolist.pasteur.fr/Subtilist/. Location/Qualifiers 1..201241 /organism="Bacillus subtilis subsp. subtilis str. 168" /mol_type="genomic DNA" /strain="168" /db_xref="taxon:224308" 67..852 /gene="yajZ" /locus_tag="BSU11350" 67..852 /gene="yajZ" /locus_tag="BSU11350" /function="unknown" /note="similar to unknown proteins" /codon_start=1 /transl_table=11 /protein_id="CAB12992.1" /db_xref="GI:2633489" /translation="ASVEQTYGWLKAGSIDLHAYIILPFGSAKKWKGILGHLOH HGFMFKIGIOIVTKLKKEGVKNKMGQGPVPIVTLVDNRRIIRL EPGSKSLAQFQMFLFLSSDLDFGSVALMTHYHHVCRGLHTKBEKDVLDTDIR MEGLAAYAYRFGSQTAETWTWTPPEOLALYEKKIAHPEDIKRDNRLFPQLLFGK GYQPRMLGYAVGNIVKYLTATASKASTADGLSIPAEITLDAML" 1044..2030 /gene="appD" /locus_tag="BSU11360" 1044..2030 /gene="appD" /locus_tag="BSU11360" /function="oligopeptide transport" /codon_start=1		/transl_table=11 /product="oligopeptide ABC transporter (ATP-binding protein)" /protein_id="CAB12993.1" /db_xref="GI:2633490" /db_xref="GOA:P42064" /db_xref="SWISS-PROT:P42064" /translation="MTLLEVNLLKTYFFRKKEPIPAVDGVDPFHISKGTVALVGESG SKGITSLSIMGLVQSOGKIMGDSIKLEDKILTSTENDYCKLRGNSVMISPMQEMPT SLNPLVTIGRTITEVLIYHKNNMKKEARQAVELLQXGVFSRAEQIMKXPRLKSGGM RORVMIALSNPKLLIADEPTALDVTTQAQVLELMKDLCOQFNTEILLITHLGVS VSEARDAILVMYCGOVENATVDLPLELPHPYTEGLTSIPVIDGEIDKLNAIKGSV PTBNLPFCGRAPRCPCAKMDKCTNQPSLLTHKSGRTVRCFLYEEGAEGS" 2027..3016 /gene="appF" /locus_tag="BSU11370" 2027..3016 /gene="appF" /locus_tag="BSU11370" /function="oligopeptide transport" /codon_start=1 /transl_table=11 /product="oligopeptide ABC transporter (ATP-binding protein)" /protein_id="CAB12994.1" /db_xref="GI:2633491" /db_xref="GOA:P42065" /db_xref="SWISS-PROT:P42065" /translation="MTAAOBTILELRDVKKYFIIRSLGFQRKVGDVKAVDGVSFSLK KGETLVIGVSGSCGCKTAGTRMIYKPTQEQLPFQODISNLSEKLRKSVKRNQIHE VPQPFAASLNPRKTLSRIIEEPENTHMVTRRENEKVELLARVGLHSPACRYPHIE FSGGQRQIGARALTINPELLIADEPVSDLVSIQPVINLMELOEENLTVLFTLS HDLSVRHSIDRVGMVIMKMMELTGKHELNDYNLPHLYTQALLSSVPVTRKRGSVKEE RIVLKGELPSPAPPKGCVFETRCPVAKPICKEQIPFEKAAPSHFVACHLYS" 3104..4735 /gene="appB" /locus_tag="BSU11380" 3104..4735 /gene="appB" /locus_tag="BSU11380" /function="oligopeptide transport" /codon_start=1 /transl_table=11 /product="oligopeptide ABC transporter /product="oligopeptide ABC transporter (permease)" /protein_id="CAB12995.1" /db_xref="GI:2633492" /db_xref="GOA:P42061" /db_xref="SWISS-PROT:P42061" /translation="MKRRKTAIAMSLLVLAIFLSACSCKSSNSAKISAGKPOQG GDLVVSGISEPTLFNSLYSTDASTDIENMLYSFLTUKDEKLNKLSIAESIKELDGG LAIFYKIKKVRFHDKELTDADVVTYSVPLSKDYKGERGSTYEMLSKSVKKGDEVY LFALKYDKGNFYNNALDSTAIIIPHILANGVPIADIENEFRKRPISGSPFKFWKWK GOVIKLEANDDYFEGRPVLDVTYKVIPDNAEAQALQAGINFNPVPAIDYKTAERF NMLKIVTDLALSUVYIGWNEKNEKPKKKVQALTTADRESIVSQLVDGDEVAIYIER ESPLSNYPKDDVPKAFYNEKAKMLAEGWNTNGDIGLDKDGKKSFSTLATNQOC NKVEDIAVVVOQLKKIGIEVKTQIWEVSALVEQMNPNDPDAMWGWSLSYTFDQC YDIFHDSOIKKGLNYVYNKADKMKDAKSIDSRKOYSKEYEIQYKXIADQPYTH LVYPNNHAMPENLEGYKHPKRDLYNIEKMWLAK" 4811..5764 /gene="appB" /locus_tag="BSU11390" 4811..5764 /gene="appB" /locus_tag="BSU11390" /function="oligopeptide transport" /codon_start=1 /transl_table=11 /product="oligopeptide ABC transporter (permease)" /protein_id="CAB12996.1" /db_xref="GI:2633493" /db_xref="GOA:P42062" /db_xref="SWISS-PROT:P42062"
--	--	---	-----------------------------------	---------------------	--	--	--

```

/translation="MAAYIIRRTLMSPILLGITILTSFVIMKAAEGDPMTLAMDPKIL
QADREQIIEYGLMDPGYQVYLKWLGNMQGDGFTSIVRKGTPVSELIMARLPNTILL
MLVSTIETKYLSPFGLVSKRYPKIDYGITFTSFIGLAIPNFPNGLLIMVLSVNL
GMFTGGVETLNTFNTFNRHLLHPAPVLATADMAGLTRYTRSNMLDLNQDYIRT
ARAKQENRVNLFKHGRLNALLPVITIFGLMIPSEFTSGSVVVEQIFTPWPLGLGKLFVDS
AFQRTIIRIMAMTVISAVLVVGVGNLIADLIYAIVDPRYEY"
5781..6692
/genes="appC"
/locus_tag="BSU11400"
5781..6692
/genes="appC"
/locus_tag="BSU11400"
/function="oligopeptide transport"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter (permease)"
/protein_id="CAB12997.1"
/db_xref="GI:2633494"
/db_xref="GOA:P42063"
/db_xref="SWISS-PROT:P42063"
/translation="MSELTQTPSPSEIRLKENISKKPTMTKIPIWEKFSKNKLAALGAV
ILFIILMSAVFAPILAPYAPQSQSLDLKYPAGLEHMGTDKFGRIIDIPSRILLYGARVS
LVGSPASVSGSILGTVLGALAGYPRGIVDAIVRVVDIVLSIPDIFLLITVITPKP
GVYDKILIFCLTGWTITLTVRGEFGLSRREYVLAAKTIGTKTHKILFSHLLPNALG
PIIVSATIKVGSVILAEASLYLFGFGIOPPIASMGNNLQDRQNFTVMIQAWWTPLPFG
LFIIMTVLCFNVFGDLRDLDPKNK"
5898..7650
/genes="yjbA"
/locus_tag="BSU11410"
5898..7650
/genes="yjbA"
/locus_tag="BSU11410"
/function="unknown"
/notes="similar to unknown proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12998.1"

```

Query Match	99.9%;	Score 1189.4;	DB 1;	Length 201241;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 1190;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy 1	ATGAATGTCTTAAACCGCGCGCAAGCCTTGCAGCGAGCGCTCTCAATGGCAAAAACAAA 60			
Db 80913	ATGAATGTGTAAACCGCGCGCAAGCCTTGCAGCGAGCGCTCTCATGGGAAAACAAA 80972			
Qy 61	CAGGATGCGTAATCATCCGTTTCCATGATGAATCGATGAGAAAGATGCGCCTGTTC 120			
Db 80973	CAGGATGCGTAATCATCCGTTTCCATGATGAATCGATGAGAAAGATGCGCCTGTTC 81032			
Qy 121	TTTGATGAAGAAAACCAAGTGTGGAGCGTTTTCTTTATGATGATGTCAAAAAGTTGTT 180			
Db 81033	TTTGATGAAGAAAACCAAGTGTGGAGCGTTTTCTTTATGATGATGTCAAAAAGTTGTT 81092			
Qy 181	GGGGATAAGAGTTGTTTCCAGTGTGCAGCGCGAGCAGACAAAGCTCTATTGGAAATTC 240			
Db 81093	GGGGATAAGAGTTGTTTCCAGTGTGCAGCGCGAGCAGACAAAGCTCTATTGGAAATTC 81152			
Qy 241	ATCATTTAAACATGACCCGCGCGAAGCATACAAAATCCGTTTCAGTCGTGAACAAGCCTTT 300			
Db 81153	ATCATTTAAACATGACCCGCGCGAAGCATACAAAATCCGTTTCAGTCGTGAACAAGCCTTT 81212			
Qy 301	ACTCGCGCGTGATGAAGCAATGGAAACCGAGAAITTCAGAAATCACAGATGACTGATT 360			
Db 81213	ACTCGCGCGTGATGAAGCAATGGAAACCGAGAAITTCAGAAATCACAGATGACTGATT 81272			
Qy 361	CAAAAATTTCAGGGCGCAGTGTGATTTGACCTTGTTCAGATTTTTCATACCCGCTTCGG 420			
Db 81273	CAAAAATTTCAGGGCGCAGTGTGATTTGACCTTGTTCAGATTTTTCATACCCGCTTCGG 81332			
Qy 421	GTTATTCGTGATATCTGAGCTGCTGGGAGTGCCCTTCAGCGCAGATGGAACAGTTTAAAGCA 480			
Db 81333	GTTATTCGTGATATCTGAGCTGCTGGGAGTGCCCTTCAGCGCAGATGGAACAGTTTAAAGCA 81392			

QY	481	TGGTCTGATCTTCTGGT	CAGTACACCGAAGATAAAGT	GAGAGCTCGA	AAAAGCCTTT	540
DB	81393	TGGTCTGATCTTCTGGT	CAGTACACCGAAGATAAAGT	GAAGAGCTG	AAAAGCCTTT	81452
QY	541	TTGGAAGACGAGATAAGT	TGTGAGGAAGAACT	TGGCGCGCTTTT	TTTTCGCGGCATCATAGAA	600
DB	81453	TTGGAAGAACGAGATAAGT	TGTGAGGAGAACT	TGGCGCGCTTTT	TTTTCGCGGCATCATAGAA	81512
QY	601	GAAAAACGAAACAAAC	CCGGAAACAGGATATATT	CTATTTTAGTGAAGCGGAGAAACA		660
DB	81513	GAAAAACGAAACAAAC	CCGGAAACAGGATATATT	CTATTTTAGTGAAGCGGAGAAACA		81572
QY	661	GCGGAGAGAGCTCTCGGT	TGAAGAGCTGATTTCGGTTT	TGCAAGCTGCTGCTGTGTGGCGCGGA		720
DB	81573	GCGGAGAGAGCTCTCGGT	TGAAGAGCTGATTTCGGTTT	TGCAAGCTGCTGCTGTGTGGCGCGGA		81632
QY	721	AATGAAACCACTACAAAC	CTGATTTTCAATTCGGATCTACAGCATAT	TTAGAAACGCCAGGC		780
DB	81633	AATGAAACCACTACAAAC	CTGATTTTCAATTCGGATCTACAGCATAT	TTAGAAACGCCAGGC		81692
QY	781	GTTTACGAGGAAGCTCGG	CAGGCCATCTGAACTGATGCCT	CAGCAGCTGGAAGAGCCCTTG		840
DB	81693	GTTTACGAGGAAGCTCGG	CAGGCCATCTGAACTGATGCCT	CAGCAGCTGGAAGAGCCCTTG		81752
QY	841	CGTTTCAGAGCGCGGCC	CCCGCTTTT	SAGCGCATTTGCCAAGCGGGATACGAGATCGGG		900
DB	81753	CGTTTCAGAGCGCGGCC	CCCGCTTTT	SAGCGCATTTGCCAAGCGGGATACGAGATCGGG		81812
QY	901	GGGCACCTGATTAAGAAGGT	GATATGTTTGGCGGTTTGGCGATCGGCATCGCAAAATCGTGAT			960
DB	81813	GGGCACCTGATTAAGAAGGT	GATATGTTTGGCGGTTTGGCGATCGGCATCGCAAAATCGTGAT			81872
QY	961	GAAGCAAAAGTTTGACACAGC	GCACATGTTTGATATCGCCGCCCATCCCAATCCGCATATT			1020
DB	81873	GAAGCAAAAGTTTGACACAGC	GCACATGTTTGATATCGCCGCCCATCCCAATCCGCATATT			81932
QY	1021	GGCTTTGGCCACGGCAT	CCATTTTGGCTTGGGGCCCGCTTGCCCGCTCTTGAGAGCAAT			1080
DB	81933	GGCTTTGGCCACGGCAT	CCATTTTGGCTTGGGGCCCGCTTGCCCGCTCTTGAGAGCAAT			81992
QY	1081	ATCGCGTTAAAGCTTTTGATT	CTGCTTTTCCCTCATATGAGTGCCTCAGTATCACTCCG			1140
DB	81993	ATCGCGTTAAAGCTTTTGATT	CTGCTTTTCCCTCATATGAGTGCCTCAGTATCACTCCG			82052
QY	1141	ATTGAAAAACAGTGTGAT	ATACGANTAAAGAGCTTCGGTGTGAAATGTAA			1191
DB	82053	ATTGAAAAACAGTGTGAT	ATACGANTAAAGAGCTTCGGTGTGAAATGTAA			82103

```

RESULT 3
AX4333775 LOCUS 966 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 2130 from Patent WO0229113.
ACCESSION AX433775
VERSION AX433775.1 GI:21658583
KEYWORDS
SOURCE Bacillus licheniformis
ORGANISM Bacillus licheniformis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1
AUTHORS Berka,R. and Clausen,I.G.
TITLE Methods for monitoring multiple gene expression
JOURNAL Patent: WO 0229113-A 2190 11-APR-2002;
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
FEATURES
source
1..966
/organism="Bacillus licheniformis"
/mol_type="unassigned DNA"
/db_xref="taxon:1402"
ORIGIN

```

Query Match 35.7%; Score 425; DB 6; Length 966;

Best Local Similarity 65.4%; Pred. No. 6.8e-112; Matches 623; Conservative 0; Mismatches 330; Indels 0; Gaps 0;	
QY	230 TTGGAAATTCATATTAACATGACCGCCGCAAGCATACAAAATCCGTTGAGTCGTCA 289
Db	7 TTGGAAATTCATATTAACATGACCGCCGCAAGCATACAAAATCCGTTGAGTCGTCA 66
QY	290 ACAAGACCTTTACTCCGCGGTGATGAAGCAATGGAAACCGAGAAATTCAGAAATCACAG 349
Db	67 ACAAGACCTTTACTCCGCGGTGATGAAGCAATGGAAACCGAGAAATTCAGAAATCACAG 126
QY	350 ATGAATCTGATTCACAAATTTACGGGCGCGAGTGTGACCTTTGTTACGATTTTCAT 409
Db	127 ATCAATTTGCTGATGTCGCGCGAGGAAAGATCGACCTTTGACAGGATTTTCAT 186
QY	410 ACCCGCTTCGGTTATTTGATATCTGAGCTGCGGAGTGCCTTCAGCGCATGGAAC 469
Db	187 ATCCGTTGCCGTAATTCGCAATTCGGAATTTGCTCGCGTTCTTTGGTGTATAGCATC 246
QY	470 AGTTTAAAGCATGGTCTGATCTTCTGTCAGTACACCGAAGGATAAAAGTGAAGAACTG 529
Db	247 ATTTCAAGAAATGGTCCGACCTGTTGTCAGCTCGCGAAGAGCGATCGCGCGAAGATG 366
QY	530 AAAAGCCCTTTTGAAGAAACGAGTATGTTGAGGAAGACTGGCGCGTTTTCGCG 589
Db	307 TGAACGAGTGAAGAAACATCAGGACCAAGCGGGAAGAGAGCTGACCGCATTCCTTTGAAA 366
QY	590 GCATCATAGAAAGCAACAAACCGAACAGGATATTTCTATTTTGTAGTGAAG 649
Db	367 AGATGATTAAGAGAAAGCAACAACTCGGCATGACCTGATTTGCTTTTGCATCAAG 426
QY	650 CGGAAGAAACAGCGAGAGCTGTCCGGTGAAGAGCTGATTCGTTTTCACCGCTGCTG 709
Db	427 CGGAACAGAGGGAAGCTGTCCCTGATGAATGCTTCCATTTTGCATCTGCTCT 486
QY	710 TGGTGCAGGAAATGAACCACTAACHAACCCTGATTTCAATGCGATGACAGCATATTAG 769
Db	487 TGATGCGGGGAATGAACGACGACGAATTTAGTTTCAATGCGGCTACAGCATTTCTCG 546
QY	770 AAACGCGAGCGTTTACGAGGAACTGCGAGCATCTGAACTGATGCTTCAGCGAGTGG 829
Db	547 AAACGCGGCGGTATGATGAGCTGCGCAGGATCTGAACTGATTCGCGAGCGAGTGG 606
QY	830 AGGAACCTTTCGCTTTACAGAGCGCGCCCGCTTTTGAAGCGCATTTGCCAAGCGGATA 889
Db	607 AGGAACCTTTCGCTTTCCGCGCCAGCGCGCATGATTTGTCGCTTCGTTTAAACAGGATA 666
QY	890 CGGAGATCGGGGCGACCTGATTAAGAAAGTGTATGTTTGGCGTTTGTGGCATCGG 949
Db	667 CCGAGATCGAGGAGTAAAGCTTGAAGAAAGGAAAGGTGTGATGCTTTCTTCTTCTG 726
QY	950 CAAATCGTGAAGCAAAAGTTTGACAGCGCACATGTTTGTATATCCGCGGCATCCCA 1009
Db	727 CCAACGTTGATGAACGAATTTTGAAGGCGGACGAAATTTGATATTCACGCGCATCGA 786
QY	1010 ATCCGATATTCGCTTTGGCCACCGCATTCATTTTCCCTTGGGCGCCCGCTTGGCCGTC 1069
Db	787 ACCCGCATATCGCATTTGGGCAAGCATCCATTTTGTGTTGGGCGCCCGCTTCGCAAGGC 846
QY	1070 TTGAAGCAATATCGGTTAAGCTTTGATTTCTGCTTTTCTTCATATGAGTGGTCTCA 1129
Db	847 TGAAGCGGCGATCGCAATTTGAGCGCTCTTAAAGCAGTATGCTTCGATGGAAGCTTG 906
QY	1130 GTATCATCTCGATTGAAGAACAGTGTGATATACGAAATTAAGAGCTTCCGTTGTG 1182
Db	907 CGGTGCGCGGATGCGGACGACGAGCATGTACGCTTGAACATTTTCGTTCTG 959
RESULT 4	
AX433767	
LOCUS	1213 bp DNA linear PAT 28-JUN-2002
DEFINITION	Sequence 2182 from Patent WO0223113.
ACCESSION	AX433767

AX433767.1 GI:21658575	
Keywords: Bacillus licheniformis	
SOURCE: Bacillus licheniformis	
ORGANISM: Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
REFERENCE: Berka, R. and Clausen, I. G.	
AUTHORS: Methods for monitoring multiple gene expression	
TITLE: Patent: WO 0223113-A 2182 11-APR-2002;	
JOURNAL: Novozymes Biotech, Inc. (US); Novozymes A/S (DK)	
FEATURES: Location/Qualifiers	
1..1213	
Source	
/organism="Bacillus licheniformis"	
/mol_type="unassigned DNA"	
/db_xref="taxon:1402"	
ORIGIN	
Query Match 20.3%; Score 242; DB 6; Length 1213;	
Best Local Similarity 53.6%; Pred. No. 1.1e-58;	
Matches 550; Conservative 0; Mismatches 470; Indels 6; Gaps 2;	
QY	75 TCGTTTCCATGTAATCGATGAGAAAGGATCGCCTGTTTCTTTGATGAGAGAAA 134
Db	92 TCGTTTCCGATTTAAACCGGATCAGAAAGAGCTCTCCTGTGCGTTTACGATGAGCG 151
QY	135 CCAAGTGTGGAGCGTTTCTTTTATGATGATGTCAAAAGAGTTGTTGGGATAAAGATT 194
Db	152 AAAATCCTTTGATATCTTTTATGATGAGACGTTCAATTCGTTTAAAACCGGAGCT 211
QY	195 GTTTTCCAGTTGATCGCGCAGCAGCAAGCTCTATTGAAATTTCCATCTTAACATGGA 254
Db	212 CTCTC-----TTCAAAACGGCGGTAATATGAA--GGAAAAAGTATATTAACAATGGA 265
QY	255 CCGCGGAGAGCATACAAAATCCGTTTCAGTCGTGAAACAAAGCCTTTACTCGCGCGTAT 314
Db	266 CCGCGGAGAGCACACAAAATGAGAGCCATCGTTAATAAAGCTTTTACGCGGAAAGCGGT 325
QY	315 GAAGCATGGAAACCGAGATTCAGAAATCAGATGAACTGATTCAMAAATTTTCAGGG 374
Db	326 GAAAGAGCTTGAACCGCATATCGAAGATGACGCTTTTATTATTAACGAGGAAACA 385
QY	375 GCGCAGTGAAGTTGACCTTTTTCACGATTTTTCATACCCGCTTCGCTTATTTGATATC 434
Db	386 GAAAGATTTGTTGATGTTGTTGAGCTTGGCTGCTCTCTTCCGCTTATTAATCATCG 445
QY	435 TGAGCTGCTGGAGTGCCTTCAGCGCAGATGAGAACAGTTTAAAGCATGTTTCTTCTTCT 494
Db	446 TGAACCTTTAGGCGTTTCCGCTGAGAGCCGCTCATGTTTAAACATTTTACAGACATCT 505
QY	495 GGTCACTACACCGAAGATAAAGTGAAGAGCTGAAAGAGCCTTTTGGAAAGACGAGA 554
Db	506 TGTGCGAGTGGCGAAGACCGCTCGCTGAAGCGCCGACGATGTACAAACGACGTGA 565
QY	555 TAAGTGTGAGGAGAACTGCGCGCTTTTTCGCGCATCATAGAGAAAGACGAAACAA 614
Db	566 AGAAGCAATCGGTTTTCGCGGATTTTAAAAACATTTATCAAGCAGCGCAAAAAGA 625
QY	615 ACCGGAACAGGATATTTCTATTTTGTAGGAGCGGAGAAACAGCGGAGAGCTGTC 674
Db	626 GCCAAAAGAGCAGCTGATTTTCGCTTTTACTGCGGCGGAAAGTTGACGGCAATCTGCTGAC 685
QY	675 CGGTGAGAGCTGATTCGCTTTTTCGACGCTGCTGCTGCGCGGAAATGAAACACCTAC 734
Db	686 AGAAGAGAGACTGCTTCTATTTTTCATCTCTTTTGTGCGAGCATGAGACGACAC 745
QY	735 AAACCTGATTTCAAAATCGGATGTACAGCATATTAGAAACGCGCGCTTTACAGGAGCT 794
Db	746 CAATTTGATCGAAACAGCGTCCGCTATCTCACAGAGATATAATCACACAGGAGCGGT 805
QY	795 GCGCAGCCATCTGAACTGATGCTCTGAGGAGAGCTTGGCTTTTCAGAGCGCC 854
Db	806 AAGACAAGATCCGTCCTCGTCCCTGTCTTTGTTGAAAGAAATGCTGCTTATTATTCGCC 865

QY 855 GGCCTGGGTTTGGAGCGCATGCTCCAAAGCGGGATACGAGATCGGGGGCACCTGATTAA 914
DB 866 CTGCGAAGCGATCGCGCGACCGCGCGAGAGAGCTTGATATCGGAGCGGTGAGGATTGC 925
QY 915 AGAAGTGATATGTTTGGCGTTTGGCGATCGGCAATCGTGATGAGCAAGCTTTGA 974
DB 926 AAAGGTTCTACAGTGTATGATGAGTGGTGGCTTCAGCGAATCGTGCAGAACTTAAAGTTGA 985
QY 975 CAGACCGGACATGTTTGATATCGCGCCCATCCCAATCGCATATTCGTTTGGCCACCG 1034
DB 986 CGATCTGTGACGCTTCAAGCTTGTATCGCAATCAACCTCATATGAGCTTCGGCTCGG 1045
QY 1035 CATCCATTTTGGCGGCGCCGCTTCCCGCTTGAAGCAAAATATCGGTTAAGTTC 1094
DB 1046 CATCCATTTTGGCGGCGCTCCCTCGCGGCTTGAAGCAAAAGTCCGCTCGATTAA 1105
QY 1095 TTGTAT 1100
DB 1106 CTGTCT 1111
RESULT 5
AE017032 290685 bp DNA linear BCT 30-APR-2003
LOCUS Bacillus anthracis str. Ames section 9 of 18 of the complete
DEFINITION genome.
ACCESSION AE017032.1 GI:30257133
VERSION AE017032.1
KEYWORDS
SOURCE
ORGANISM
Bacillus anthracis str. Ames
Bacillus anthracis str. Ames
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.
1 (bases 1 to 290685)
Read, T., Peterson, S., Tourasse, N., Baillie, L., Paulsen, I.,
Nelson, K., Tettelin, H., Fouts, D., Eisen, J., Gill, S., Holtzapple, E.,
Okstad, O., Helgason, E., Rilstone, J., Wu, M., Kolonay, J., Beanan, M.,
Dodson, R., Brinkac, L., Gwinn, M., DeBoy, R., Madupu, R., Daugherty, S.,
Darkin, A., Haft, D., Nelson, W., Peterson, J., Pop, M., Khouri, H.,
Radune, D., Benton, J., Mahamoud, Y., Jiang, L., Hance, I., Weidman, J.,
Berry, K., Plaut, R., Wolf, A., Watkins, K., Nierman, W., Hazen, A.,
Cline, R., Redmond, C., Thwaite, J., White, O., Salzberg, S.,
Thomason, B., Friedlander, A., Koehler, T., Hanna, P., Kolsto, A.-B. and
Fraser, C.
The genome sequence of Bacillus anthracis Ames and comparison to
closely related bacteria
Nature 423 (6935), 81-86 (2003)
22608414
12721629
REFERENCE 2 (bases 1 to 290685)
Read, T., Peterson, S., Tourasse, N., Baillie, L., Paulsen, I.,
Nelson, K., Tettelin, H., Fouts, D., Eisen, J., Gill, S., Holtzapple, E.,
Okstad, O., Helgason, E., Rilstone, J., Wu, M., Kolonay, J., Beanan, M.,
Dodson, R., Brinkac, L., Gwinn, M., DeBoy, R., Madupu, R., Daugherty, S.,
Darkin, A., Haft, D., Nelson, W., Peterson, J., Pop, M., Khouri, H.,
Radune, D., Benton, J., Mahamoud, Y., Jiang, L., Hance, I., Weidman, J.,
Berry, K., Plaut, R., Wolf, A., Watkins, K., Nierman, W., Hazen, A.,
Cline, R., Redmond, C., Thwaite, J., White, O., Salzberg, S.,
Thomason, B., Friedlander, A., Koehler, T., Hanna, P., Kolsto, A.-B. and
Fraser, C.
Direct Submission
Submitted (26-MAR-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
FEATURES
source
1. 290685
/organism="Bacillus anthracis str. Ames"
/mol_type="genomic DNA"
/strain="Ames"
/db_xref="taxon:198094"
213..2344
/locus_tag="BA2506"
/note="this region contains an authentic frame shift and
is not the result of a sequencing artifact;"
gene
213..2344
/locus_tag="BA2506"

penicillin-binding protein, authentic frameshift; similar
to GB:100264, GB:X06989, GB:Y24546, GB:M24547, GB:M34862,
GB:M34863, GB:M34864, GB:M34865, GB:M34866, GB:M34867,
GB:M34868, GB:M34869, GB:M34870, GB:M34871, GB:M34872,
GB:M34873, GB:M34874, GB:M34875, GB:M34876, GB:M34877, GB:M34878,
GB:M34879, GB:M34880, GB:M34881, GB:M34882, GB:M34883,
GB:M34884, GB:M34885, GB:M34886, GB:M34887, GB:M34888,
GB:M34889, GB:M34890, GB:M34891, GB:M34892, GB:M34893,
GB:M34894, GB:M34895, GB:M34896, GB:M34897, GB:M34898,
GB:M34899, GB:M34900, GB:M34901, GB:M34902, GB:M34903,
GB:M34904, GB:M34905, GB:M34906, GB:M34907, GB:M34908,
GB:M34909, GB:M34910, GB:M34911, GB:M34912, GB:M34913,
GB:M34914, GB:M34915, GB:M34916, GB:M34917, GB:M34918,
GB:M34919, GB:M34920, GB:M34921, GB:M34922, GB:M34923,
GB:M34924, GB:M34925, GB:M34926, GB:M34927, GB:M34928,
GB:M34929, GB:M34930, GB:M34931, GB:M34932, GB:M34933,
GB:M34934, GB:M34935, GB:M34936, GB:M34937, GB:M34938,
GB:M34939, GB:M34940, GB:M34941, GB:M34942, GB:M34943,
GB:M34944, GB:M34945, GB:M34946, GB:M34947, GB:M34948,
GB:M34949, GB:M34950, GB:M34951, GB:M34952, GB:M34953,
GB:M34954, GB:M34955, GB:M34956, GB:M34957, GB:M34958,
GB:M34959, GB:M34960, GB:M34961, GB:M34962, GB:M34963,
GB:M34964, GB:M34965, GB:M34966, GB:M34967, GB:M34968,
GB:M34969, GB:M34970, GB:M34971, GB:M34972, GB:M34973,
GB:M34974, GB:M34975, GB:M34976, GB:M34977, GB:M34978,
GB:M34979, GB:M34980, GB:M34981, GB:M34982, GB:M34983,
GB:M34984, GB:M34985, GB:M34986, GB:M34987, GB:M34988,
GB:M34989, GB:M34990, GB:M34991, GB:M34992, GB:M34993,
GB:M34994, GB:M34995, GB:M34996, GB:M34997, GB:M34998,
GB:M34999, GB:M35000, GB:M35001, GB:M35002, GB:M35003,
GB:M35004, GB:M35005, GB:M35006, GB:M35007, GB:M35008,
GB:M35009, GB:M35010, GB:M35011, GB:M35012, GB:M35013,
GB:M35014, GB:M35015, GB:M35016, GB:M35017, GB:M35018,
GB:M35019, GB:M35020, GB:M35021, GB:M35022, GB:M35023,
GB:M35024, GB:M35025, GB:M35026, GB:M35027, GB:M35028,
GB:M35029, GB:M35030, GB:M35031, GB:M35032, GB:M35033,
GB:M35034, GB:M35035, GB:M35036, GB:M35037, GB:M35038,
GB:M35039, GB:M35040, GB:M35041, GB:M35042, GB:M35043,
GB:M35044, GB:M35045, GB:M35046, GB:M35047, GB:M35048,
GB:M35049, GB:M35050, GB:M35051, GB:M35052, GB:M35053,
GB:M35054, GB:M35055, GB:M35056, GB:M35057, GB:M35058,
GB:M35059, GB:M35060, GB:M35061, GB:M35062, GB:M35063,
GB:M35064, GB:M35065, GB:M35066, GB:M35067, GB:M35068,
GB:M35069, GB:M35070, GB:M35071, GB:M35072, GB:M35073,
GB:M35074, GB:M35075, GB:M35076, GB:M35077, GB:M35078,
GB:M35079, GB:M35080, GB:M35081, GB:M35082, GB:M35083,
GB:M35084, GB:M35085, GB:M35086, GB:M35087, GB:M35088,
GB:M35089, GB:M35090, GB:M35091, GB:M35092, GB:M35093,
GB:M35094, GB:M35095, GB:M35096, GB:M35097, GB:M35098,
GB:M35099, GB:M35100, GB:M35101, GB:M35102, GB:M35103,
GB:M35104, GB:M35105, GB:M35106, GB:M35107, GB:M35108,
GB:M35109, GB:M35110, GB:M35111, GB:M35112, GB:M35113,
GB:M35114, GB:M35115, GB:M35116, GB:M35117, GB:M35118,
GB:M35119, GB:M35120, GB:M35121, GB:M35122, GB:M35123,
GB:M35124, GB:M35125, GB:M35126, GB:M35127, GB:M35128,
GB:M35129, GB:M35130, GB:M35131, GB:M35132, GB:M35133,
GB:M35134, GB:M35135, GB:M35136, GB:M35137, GB:M35138,
GB:M35139, GB:M35140, GB:M35141, GB:M35142, GB:M35143,
GB:M35144, GB:M35145, GB:M35146, GB:M35147, GB:M35148,
GB:M35149, GB:M35150, GB:M35151, GB:M35152, GB:M35153,
GB:M35154, GB:M35155, GB:M35156, GB:M35157, GB:M35158,
GB:M35159, GB:M35160, GB:M35161, GB:M35162, GB:M35163,
GB:M35164, GB:M35165, GB:M35166, GB:M35167, GB:M35168,
GB:M35169, GB:M35170, GB:M35171, GB:M35172, GB:M35173,
GB:M35174, GB:M35175, GB:M35176, GB:M35177, GB:M35178,
GB:M35179, GB:M35180, GB:M35181, GB:M35182, GB:M35183,
GB:M35184, GB:M35185, GB:M35186, GB:M35187, GB:M35188,
GB:M35189, GB:M35190, GB:M35191, GB:M35192, GB:M35193,
GB:M35194, GB:M35195, GB:M35196, GB:M35197, GB:M35198,
GB:M35199, GB:M35200, GB:M35201, GB:M35202, GB:M35203,
GB:M35204, GB:M35205, GB:M35206, GB:M35207, GB:M35208,
GB:M35209, GB:M35210, GB:M35211, GB:M35212, GB:M35213,
GB:M35214, GB:M35215, GB:M35216, GB:M35217, GB:M35218,
GB:M35219, GB:M35220, GB:M35221, GB:M35222, GB:M35223,
GB:M35224, GB:M35225, GB:M35226, GB:M35227, GB:M35228,
GB:M35229, GB:M35230, GB:M35231, GB:M35232, GB:M35233,
GB:M35234, GB:M35235, GB:M35236, GB:M35237, GB:M35238,
GB:M35239, GB:M35240, GB:M35241, GB:M35242, GB:M35243,
GB:M35244, GB:M35245, GB:M35246, GB:M35247, GB:M35248,
GB:M35249, GB:M35250, GB:M35251, GB:M35252, GB:M35253,
GB:M35254, GB:M35255, GB:M35256, GB:M35257, GB:M35258,
GB:M35259, GB:M35260, GB:M35261, GB:M35262, GB:M35263,
GB:M35264, GB:M35265, GB:M35266, GB:M35267, GB:M35268,
GB:M35269, GB:M35270, GB:M35271, GB:M35272, GB:M35273,
GB:M35274, GB:M35275, GB:M35276, GB:M35277, GB:M35278,
GB:M35279, GB:M35280, GB:M35281, GB:M35282, GB:M35283,
GB:M35284, GB:M35285, GB:M35286, GB:M35287, GB:M35288,
GB:M35289, GB:M35290, GB:M35291, GB:M35292, GB:M35293,
GB:M35294, GB:M35295, GB:M35296, GB:M35297, GB:M35298,
GB:M35299, GB:M35300, GB:M35301, GB:M35302, GB:M35303,
GB:M35304, GB:M35305, GB:M35306, GB:M35307, GB:M35308,
GB:M35309, GB:M35310, GB:M35311, GB:M35312, GB:M35313,
GB:M35314, GB:M35315, GB:M35316, GB:M35317, GB:M35318,
GB:M35319, GB:M35320, GB:M35321, GB:M35322, GB:M35323,
GB:M35324, GB:M35325, GB:M35326, GB:M35327, GB:M35328,
GB:M35329, GB:M35330, GB:M35331, GB:M35332, GB:M35333,
GB:M35334, GB:M35335, GB:M35336, GB:M35337, GB:M35338,
GB:M35339, GB:M35340, GB:M35341, GB:M35342, GB:M35343,
GB:M35344, GB:M35345, GB:M35346, GB:M35347, GB:M35348,
GB:M35349, GB:M35350, GB:M35351, GB:M35352, GB:M35353,
GB:M35354, GB:M35355, GB:M35356, GB:M35357, GB:M35358,
GB:M35359, GB:M35360, GB:M35361, GB:M35362, GB:M35363,
GB:M35364, GB:M35365, GB:M35366, GB:M35367, GB:M35368,
GB:M35369, GB:M35370, GB:M35371, GB:M35372, GB:M35373,
GB:M35374, GB:M35375, GB:M35376, GB:M35377, GB:M35378,
GB:M35379, GB:M35380, GB:M35381, GB:M35382, GB:M35383,
GB:M35384, GB:M35385, GB:M35386, GB:M35387, GB:M35388,
GB:M35389, GB:M35390, GB:M35391, GB:M35392, GB:M35393,
GB:M35394, GB:M35395, GB:M35396, GB:M35397, GB:M35398,
GB:M35399, GB:M35400, GB:M35401, GB:M35402, GB:M35403,
GB:M35404, GB:M35405, GB:M35406, GB:M35407, GB:M35408,
GB:M35409, GB:M35410, GB:M35411, GB:M35412, GB:M35413,
GB:M35414, GB:M35415, GB:M35416, GB:M35417, GB:M35418,
GB:M35419, GB:M35420, GB:M35421, GB:M35422, GB:M35423,
GB:M35424, GB:M35425, GB:M35426, GB:M35427, GB:M35428,
GB:M35429, GB:M35430, GB:M35431, GB:M35432, GB:M35433,
GB:M35434, GB:M35435, GB:M35436, GB:M35437, GB:M35438,
GB:M35439, GB:M35440, GB:M35441, GB:M35442, GB:M35443,
GB:M35444, GB:M35445, GB:M35446, GB:M35447, GB:M35448,
GB:M35449, GB:M35450, GB:M35451, GB:M35452, GB:M35453,
GB:M35454, GB:M35455, GB:M35456, GB:M35457, GB:M35458,
GB:M35459, GB:M35460, GB:M35461, GB:M35462, GB:M35463,
GB:M35464, GB:M35465, GB:M35466, GB:M35467, GB:M35468,
GB:M35469, GB:M35470, GB:M35471, GB:M35472, GB:M35473,
GB:M35474, GB:M35475, GB:M35476, GB:M35477, GB:M35478,
GB:M35479, GB:M35480, GB:M35481, GB:M35482, GB:M35483,
GB:M35484, GB:M35485, GB:M35486, GB:M35487, GB:M35488,
GB:M35489, GB:M35490, GB:M35491, GB:M35492, GB:M35493,
GB:M35494, GB:M35495, GB:M35496, GB:M35497, GB:M35498,
GB:M35499, GB:M35500, GB:M35501, GB:M35502, GB:M35503,
GB:M35504, GB:M35505, GB:M35506, GB:M35507, GB:M35508,
GB:M35509, GB:M35510, GB:M35511, GB:M35512, GB:M35513,
GB:M35514, GB:M35515, GB:M35516, GB:M35517, GB:M35518,
GB:M35519, GB:M35520, GB:M35521, GB:M35522, GB:M35523,
GB:M35524, GB:M35525, GB:M35526, GB:M35527, GB:M35528,
GB:M35529, GB:M35530, GB:M35531, GB:M35532, GB:M35533,
GB:M35534, GB:M35535, GB:M35536, GB:M35537, GB:M35538,
GB:M35539, GB:M35540, GB:M35541, GB:M35542, GB:M35543,
GB:M35544, GB:M35545, GB:M35546, GB:M35547, GB:M35548,
GB:M35549, GB:M35550, GB:M35551, GB:M35552, GB:M35553,
GB:M35554, GB:M35555, GB:M35556, GB:M35557, GB:M35558,
GB:M35559, GB:M35560, GB:M35561, GB:M35562, GB:M35563,
GB:M35564, GB:M35565, GB:M35566, GB:M35567, GB:M35568,
GB:M35569, GB:M35570, GB:M35571, GB:M35572, GB:M35573,
GB:M35574, GB:M35575, GB:M35576, GB:M35577, GB:M35578,
GB:M35579, GB:M35580, GB:M35581, GB:M35582, GB:M35583,
GB:M35584, GB:M35585, GB:M35586, GB:M35587, GB:M35588,
GB:M35589, GB:M35590, GB:M35591, GB:M35592, GB:M35593,
GB:M35594, GB:M35595, GB:M35596, GB:M35597, GB:M35598,
GB:M35599, GB:M35600, GB:M35601, GB:M35602, GB:M35603,
GB:M35604, GB:M35605, GB:M35606, GB:M35607, GB:M35608,
GB:M35609, GB:M35610, GB:M35611, GB:M35612, GB:M35613,
GB:M35614, GB:M35615, GB:M35616, GB:M35617, GB:M35618,
GB:M35619, GB:M35620, GB:M35621, GB:M35622, GB:M35623,
GB:M35624, GB:M35625, GB:M35626, GB:M35627, GB:M35628,
GB:M35629, GB:M35630, GB:M35631, GB:M35632, GB:M35633,
GB:M35634, GB:M35635, GB:M35636, GB:M35637, GB:M35638,
GB:M35639, GB:M35640, GB:M35641, GB:M35642, GB:M35643,
GB:M35644, GB:M35645, GB:M35646, GB:M35647, GB:M35648,
GB:M35649, GB:M35650, GB:M35651, GB:M35652, GB:M35653,
GB:M35654, GB:M35655, GB:M35656, GB:M35657, GB:M35658,
GB:M35659, GB:M35660, GB:M35661, GB:M35662, GB:M35663,
GB:M35664, GB:M35665, GB:M35666, GB:M35667, GB:M35668,
GB:M35669, GB:M35670, GB:M35671, GB:M35672, GB:M35673,
GB:M35674, GB:M35675, GB:M35676, GB:M35677, GB:M35678,
GB:M35679, GB:M35680, GB:M35681, GB:M35682, GB:M35683,
GB:M35684, GB:M35685, GB:M35686, GB:M35687, GB:M35688,
GB:M35689, GB:M35690, GB:M35691, GB:M35692, GB:M35693,
GB:M35694, GB:M35695, GB:M35696, GB:M35697, GB:M35698,
GB:M35699, GB:M35700, GB:M35701, GB:M35702, GB:M35703,
GB:M35704, GB:M35705, GB:M35706, GB:M35707, GB:M35708,
GB:M35709, GB:M35710, GB:M35711, GB:M35712, GB:M35713,
GB:M35714, GB:M35715, GB:M35716, GB:M35717, GB:M35718,
GB:M35719, GB:M35720, GB:M35721, GB:M35722, GB:M35723,
GB:M35724, GB:M35725, GB:M35726, GB:M35727, GB:M35728,
GB:M35729, GB:M35730, GB:M35731, GB:M35732, GB:M35733,
GB:M35734, GB:M35735, GB:M35736, GB:M35737, GB:M35738,
GB:M35739, GB:M35740, GB:M35741, GB:M35742, GB:M35743,
GB:M35744, GB:M35745, GB:M35746, GB:M35747, GB:M35748,
GB:M35749, GB:M35750, GB:M35751, GB:M35752, GB:M35753,
GB:M35754, GB:M35755, GB:M35756, GB:M35757, GB:M35758,
GB:M35759, GB:M35760, GB:M35761, GB:M35762, GB:M35763,
GB:M35764, GB:M35765, GB:M35766, GB:M35767, GB:M35768,
GB:M35769, GB:M35770, GB:M35771, GB:M35772, GB:M35773,
GB:M35774, GB:M35775, GB:M35776, GB:M35777, GB:M35778,
GB:M35779, GB:M35780, GB:M35781, GB:M35782, GB:M35783,
GB:M35784, GB:M35785, GB:M35786, GB:M35787, GB:M35788,
GB:M35789, GB:M35790, GB:M35791, GB:M35792, GB:M35793,
GB:M35794, GB:M35795, GB:M35796, GB:M35797, GB:M35798,
GB:M35799, GB:M35800, GB:M35801, GB:M35802, GB:M35803,
GB:M35804, GB:M35805, GB:M35806, GB:M35807, GB:M35808,
GB:M35809, GB:M35810, GB:M35811, GB:M35812, GB:M35813,
GB:M35814, GB:M35815, GB:M35816, GB:M35817, GB:M35818,
GB:M35819, GB:M35820, GB:M35821, GB:M35822, GB:M35823,
GB:M35824, GB:M35825, GB:M35826, GB:M35827, GB:M35828,
GB:M35829, GB:M35830, GB:M35831, GB:M35832, GB:M35833,
GB:M35834, GB:M35835, GB:M35836, GB:M35837, GB:M35838,
GB:M35839, GB:M35840, GB:M35841, GB:M35842, GB:M35843,
GB:M35844, GB:M35845, GB:M35846, GB:M35847, GB:M35848,
GB:M35849, GB:M35850, GB:M35851, GB:M35852, GB:M35853,
GB:M35854, GB:M35855, GB:M35856, GB:M35857, GB:M35858,
GB:M35859, GB:M35860, GB:M35861, GB:M35862, GB:M35863,
GB:M35864, GB:M35865, GB:M35866, GB:M35867, GB:M35868,
GB:M35869, GB:M35870, GB:M35871, GB:M35872, GB:M35873,
GB:M35874, GB:M35875, GB:M35876, GB:M35877, GB:M35878,
GB:M35879, GB:M35880, GB:M35881, GB:M35882, GB:M35883,
GB:M35884, GB:M35885, GB:M35886, GB:M35887, GB:M35888,
GB:M35889, GB:M35890, GB:M35891, GB:M35892, GB:M35893,
GB:M35894, GB:M35895, GB:M35896, GB:M35897, GB:M35898,
GB:M35899, GB:M35900, GB:M35901, GB:M35902, GB:M35903,
GB:M35904, GB:M35905, GB:M35906, GB:M35907, GB:M35908,
GB:M35909, GB:M35910, GB:M35911, GB:M35912, GB:M35913,
GB:M35914, GB:M35915, GB:M35916, GB:M35917, GB:M35918,
GB:M35919, GB:M35920, GB:M35921, GB:M35922, GB:M35923,
GB:M35924, GB:M35925, GB:M35926, GB:M35927, GB:M35928,
GB:M35929, GB:M35930, GB:M35931, GB:M35932, GB:M35933,
GB:M35934, GB:M35935, GB:M35936, GB:M35937, GB:M35938,
GB:M35939, GB:M35940, GB:M35941, GB:M35942, GB:M35943,
GB:M35944, GB:M35945, GB:M35946, GB:M35947, GB:M35948,
GB:M35949, GB:M35950, GB:M35951, GB:M35952, GB:M35953,
GB:M35954, GB:M35955, GB:M35956, GB:M35957, GB:M35958,
GB:M35959, GB:M35960, GB:M35961, GB:M35962, GB:M35963,
GB:M35964, GB:M35965, GB:M35966, GB:M35967, GB:M35968,
GB:M35969, GB:M35970, GB:M35971, GB:M35972, GB:M35973,
GB:M35974, GB:M35975, GB:M35976, GB:M35977, GB:M35978,
GB:M35979, GB:M35980, GB:M35981, GB:M35982, GB:M35983,
GB:M35984, GB:M35985, GB:M35986, GB:M35987, GB:M35988,
GB:M35989, GB:M35990, GB:M35991, GB:M35992, GB:M35993,
GB:M35994, GB:M35995, GB:M35996, GB:M35997, GB:M35998,
GB:M35999, GB:M36000, GB:M36001, GB:M36002, GB:M36003,
GB:M36004, GB:M36005, GB:M36006, GB:M36007, GB:M36008,
GB:M36009, GB:M36010, GB:M36011, GB:M36012, GB:M36013,
GB:M36014, GB:M36015, GB:M36016, GB:M36017, GB:M36018,
GB:M36019, GB:M36020, GB:M36021, GB:M36022, GB:M36023,
GB:M36024, GB:M36025, GB:M36026, GB:M36027, GB:M36028,
GB:M36029, GB:M36030, GB:M36031, GB:M36032, GB:M36033,
GB:M36034, GB:M36035, GB:M36036, GB:M36037, GB:M36038,
GB:M36039, GB:M36040, GB:M36041, GB:M36042, GB:M36043,
GB:M36044, GB:M36045, GB:M36046, GB:M36047, GB:M36048,
GB:M36049, GB:M36050, GB:M36051, GB:M36052, GB:M36053,
GB:M36054, GB:M36055, GB:M36056, GB:M36057, GB:M36058,
GB:M36059, GB:M36060, GB:M36061, GB:M36062, GB:M36063,
GB:M36064, GB:M36065, GB:M36066, GB:M36067, GB:M36068,
GB:M36069, GB:M36070, GB:M36071, GB:M36072, GB:M36073,
GB:M36074, GB:M36075, GB:M36076, GB:M36077, GB:M36078,
GB:M36079, GB:M36080, GB:M36081, GB:M36082, GB:M36083,
GB:M36084, GB:M36085, GB:M36086, GB:M36087, GB:M36088,
GB:M36089, GB:M36090, GB:M36091, GB:M36092, GB:M36093,
GB:M36094, GB:M36095, GB:M36096, GB:M36097, GB:M36098,
GB:M36099, GB:M36100, GB:M36101, GB:M36102, GB:M36103,
GB:M36104, GB:M36105, GB:M36106, GB:M36107, GB:M36108,
GB:M36109, GB:M36110, GB:M36111, GB:M36112, GB:M36113,
GB:M36114, GB:M36115, GB:M36116, GB:M36117, GB:M36118,
GB:M36119, GB:M36120, GB:M36121, GB:M36122, GB:M36123,
GB:M36124, GB:M36125, GB:M36126, GB:M36127, GB:M36128,
GB:M36129, GB:M36130, GB:M36131, GB:M36132, GB:M36133,
GB:M36134, GB:M36135, GB:M36136, GB:M36137, GB:M36138,
GB:M36139, GB:M36140, GB:M36141, GB:M36142, GB:M36143,
GB:M36144, GB:M36145, GB:M3

```

/notes="similar to SP:O05390, SP:P37134, GB:X62116, and
PID:580779; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="alpha-ketoglutarate permease"
/protein_id="AAP26368.1"
/db_xref="GI:30257137"
/translation="MNALISNKTMDRIGIPNSLTWGYIGIIVMIGDGLGEOGMLSPY
LVEKGLTLEHAAFLTYGTYVSASSWFSGVFMGMGRKVMYTFGLVSPILGSIQFIS
IGQHMYPVLICITLARGGYPLEPAYSPLVWSYSPQOMLSRAGWFWFVFOGLS
VIGAFYSYXVGVIGIVLWSALIFVVGGLFSLVFNKDKFKAQTSKANSSELKLS
ITTAFFPKVIGKIVIRKINSAAQGFVFLPYMKYNTFTWLOIWLQWLFVNMV
FNIFGLVGVKGFINTIKMFGVCGGIVTLALYVPMVGNHNYWALIFVACCYGATL
AGVPLTALVPSLSPENKRAMSVNLGSLSAFVGLVWTFATIGPLGVGVGMVIFAG
LXFFGAEZLHFLTIPTKENEMQDLNLTSGBQFSV"
/locus_tag="BA2511"
/locus_tag="BA2511"
/locus_tag="BA2511"
/notes="similar to SP:P37134, GB:X62116, and PID:580779;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, NAD-binding"
/protein_id="AAP26369.1"
/db_xref="GI:30257138"
/translation="MNVLKIGIGAGRIGKLVNQLMPOVKIKAVSDVVVIGLKKW
AQDKGSLTNNYQDLADPEIDAVFICSPYTHAQILKEAALAKHIFCEKPVFSV
BETLEALVEVQGVQGVQVQVDFNFRKYDIFQOGEVQGHILKITSRDPQPS
TEYVSSGLPMIMHDFDMARYMNSVEVEFYAGFTLLDPSIQEVDHVDIAIVTL
KFANGALVINDSQAQVGYQVVEFGEKGAADNCCPTTQVQSKTEGIVKDKPLY
FFLERYTCAYIEEVTPQTSIIIEQAVICSGNDGLQAEIRIAKAKESLLTKPKVQIEH
KOPALNC"
/locus_tag="BA2512"
/locus_tag="BA2512"
/locus_tag="BA2512"
/notes="similar to SP:P4212832, and SP:P42414; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="iolC protein, putative"
/protein_id="AAP26370.1"
/db_xref="GI:30257139"
/translation="MNPLIFKRNRPDLIAVGRICVDLNNETORPMEETRTFTKYVG
GSPANIAIGARLGLQTFIGKVDOMGRFTYGLKDKNKLNTDQIPIDCTGAVTGLA
FTEIKSPDCSLMTRDNVADLNLOPTEVSEDIYKQSKALLISGITALAKSPSREAVEL
ALEYARKHDVVFFVDYRPYPTWQSEATVAYINLAEEKSDVILITREEFDMMKILN
YEKNDQVTAERWFSSHAKIVVIRKGGDSIAYTRDQSHRGSIIPKTKVLPFGAGDS
VASAFIYGLMOGLEIPOAMRLGGASIVISKHSCSDAMPTRAEISAFMETAEELV"
/locus_tag="BA2513"
/locus_tag="BA2513"
/locus_tag="BA2513"
/notes="similar to SP:P42412; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="methylmalonic acid semialdehyde dehydrogenase"
/protein_id="AAP26371.1"
/db_xref="GI:30257140"
/translation="MTVQTAQIVKQNYIGGEVWESISTKMEAVTNPATGBVIAQVPLST
KUDVQVLAANEAPKSNKTAVPARILFKYQQLLVNWEELAKLITTEKSKNE
AYGEVLRGLECFEFPAGATPLMWGQPLDIAIGIESCMYRPIPLVIGGITPFPNPMV
PCWMPPLATACCTNFKPESRTPPLQALAEAEAGLPKGVNIVNAGHGVNGLL
EHLKVKAI SFVGSQVAYEYVYKKNLXRVQALAKAKNHSIVLNDANIELATKQIIS
AFAFGAGERCMASVYVEELADQLVERLVANKEIVLNGLEDVDFLGPVIRNKH
ERTIGYIDSGVQGLATVRDGEDTAVKAGYFVGPTIFDHVTKEMKIWDQRIEPAVL
SIVRVKSLDEALIEIANESPANGACIYTDSGASVYRQPRETIESGMLGVNVVPAPNAF
FFPSGWNKDSFYGLDGHANGTDGVEFFYTRKKMLTSRWEK"
10997..12931

```

```

/locus_tag="BA2514"
10997..12931
/locus_tag="BA2514"
/notes="similar to GP:15980566, and SP:P42415; identified
Query Match 19.7%; Score 234.4; DB 1; Length 290685;
Best Local Similarity 54.5%; Pred. No. 5.8e-56;
Matches 597; Conservative 0; Mismatches 466; Indels 33; Gaps 5;
QY 46 AATCGGAAAAACAAACAGGATCGGTATCATCCGTTTCCATCGTATGTAATCGATGAGAAAG 105
DB 123554 AAGCTGAAAACGAAGAGAAATTTGGGATCCGTATGATGATGATCAATTTATGAGGAT 123613
QY 106 GATCGGCTGTTTCTTTGATGAGAAACCAAGTGTGGAGGTTTTTCTTTTATGATGAT 165
DB 123614 AATCACCAGTACATTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 123673
QY 166 GTCAAAAAGTGTGTGGGATTAAGAGTTGTTTCCAGTTGCATGCCGAGCAGCAAGC 225
DB 123674 GTAAATCGAGTCTTATCGGATTTATAGCTTATTTTCAAGTAGAAGGAAGCAATTTT 123733
QY 226 TCTATTGGAAATTCATCAATTAACAT-----GGACCCGCGGAGGATACA 270
DB 123734 GCATCCACCTTTAGAACTAGGATTAATTAATCTCTACAGATCCACCAACATCGC 123793
QY 271 AAAATCCGTTGATCGTGAAACAAAGCCTTTACTCCGGCGGTGATGAAGCAATGGGAACG 330
DB 123794 AACGTACGTTCTCTGTTTCTTAAAGCAATTTACTCCAGAAGTTTAAACAATGGAAACCT 123853
QY 331 AGAATTCGAAGAAATCACAGATGAACCTGATTCAAAAATTTTCAGGGCGGAGTCAAGTTGAC 390
DB 123854 CGAATACAGTCTATCGCAATGAACCTGTAAAGATTTTGAAGAACTGATGATGATGAT 123913
QY 391 CTGTGTTACGATTTTTCATACCCGCTTCGGGTTATTTGATATCTAGCTGCTGGAGTG 450
DB 123914 ATCGTTGACAGTTTCTGCTCTTCTTACCTGTTACCGTCATATCGATTTATTAGAGTG 123973
QY 451 CTTTCAGCGCAGATGGAAAGAGTTTAAAGATCGGTCTGATCTTCTGTCAGTACACCGAAG 510
DB 123974 CCAACAACCTGACCGTAAAGATTTAAAGCATGCTGCTGATTTCT-----ATTATG 124024
QY 511 GATAAAGTGAAGAGCTGAAAGAGCCTTTTGGAGAGACGAGATAAGTGTGAGGAGAA 570
DB 124025 CGGTATGTAAGAAAGAAAGTTTAAATGATCTGATGAGAAAGAGGATAGCATTATGAA 124084
QY 571 CTGGCGCGGTTTTTTCGCGGCATCATAGAAAGAAAGCGAAACAAACCGGAACAGGATATT 630
DB 124085 TTTAAGGCATATCTTCTACCGATCGTTACAGAAAGAGATATCATTTAACCGATGATATT 124144
QY 631 ATTTCTATTTTGTAGTGAAGCGGAGAAACAGAGGAGAGAGCTGTCCTCGTGAAGAGCTGATT 690
DB 124145 ATCTCTGATTTAATACGAGCTGAATATGAAAGGTGAAGATTTAACTGATGAAGAAATTTG 124204
QY 691 CGGTTTTCACGCTGCTGCTGCTGGCGGAAATGAACCACTACAAACCTGATTTCAAT 750
DB 124205 ACTTTTCTTTAGGCTTATTAGCCGCGGTATTAAGAAACCTTCAATTTTATCATTAAT 124264
QY 751 GCGATGTACAGCATATTA---GAAACGCGAGCGGTTTACGAGGAACCTGCGCAGCCATCCT 807
DB 124265 AGTTTTTACGTCTTTTATAGTGACTACCTCGCAACATATAAGAAAGTAAGAGAAAAACCT 124324
QY 808 GAATGATGCCCTCAGGAGTGGAGGAGCTTGGCTTTCAGAGCGCGCGCGCGCTTTTG 867
DB 124325 AAATTAATTTTCAAGACGGTTGAGGAAGTATTACGCTATCGCTTTCGGTTACATTAGCT 124384
QY 868 AGCGCATTTGCCAAGCGGATACGGAGATCGGGGCGCACTGATTAAGAAAGGTGATG 927
DB 124385 AGGAGAAATTACAGA---GGACACCAATATATTTTGGACCTTTATGAAGAAAGGTGATG 124441
QY 928 GTTTTCGGCGTTGTGCGCATCGGCAATCGTGATGAAGCAAGTTTGCAGACCGCACATG 987
DB 124442 GTTGTTCGGTGGGTTAGTCGAGCAATTTAGATGAGAAAAAGTTTTTCAAGCATCTAAG 124501
10997..12931

```

QY	988	TTTGATATCCGCGCATCCCAAT---CCGATATTGCGTTTGGCCACCGCATCCATTTT	1044
DB	124502	TTTATATACATCGATAGGAATGAAGACATTTAAACCTTTGGTAAGGCGCTCACITTT	124561
QY	1045	TGCTTTGGGCGCGCGTTCGCGCTTCTGAAGCAATATCGGTTAACTGTTTGAATTTCT	1104
DB	124562	TGCTTTAGGCGCACCACTTGGCGTTTGAAGCTGAGATTGCATTAACTACCTTTATAAT	124621
QY	1105	GCTTTTCCTCATATGG	1120
DB	124622	GCTTTTGAAGAAGATAG	124637
RESULT 6			
BACKUP/C			
LOCUS		3187 bp DNA linear BCT 26-APR-1993	
DEFINITION		B.subtilis rtp gene, complete cds and proC gene (put.), 5' end.	
ACCESSION		M24523.1 M36988	
VERSION		M24523.1 GI:143477	
KEYWORDS		DNA-binding protein; proC gene; replication terminator protein; terC gene.	
SOURCE		Bacillus subtilis	
ORGANISM		Bacillus subtilis	
REFERENCE		Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
AUTHORS		Lewis, P.J. and Wake, R.G.	
TITLE		DNA and protein sequence conservation at the replication terminus	
JOURNAL		In Bacillus subtilis 168 and W23	
MEDLINE		J. Bacteriol. 171 (3), 1402-1408 (1989)	
PUBMED		89155440	
REFERENCE		2493444	
AUTHORS		2 (sites)	
TITLE		Ahn, K. and Wake, R.G.	
JOURNAL		A unique open reading frame adjacent to the replication terminus of	
MEDLINE		the Bacillus subtilis W23 chromosome compared with Bacillus	
PUBMED		subtilis 168	
REFERENCE		Unpublished (1990)	
AUTHORS		3 (sites)	
TITLE		Ahn, K.S. and Wake, R.G.	
JOURNAL		Variations and coding features of the sequence spanning the	
MEDLINE		replication terminus of Bacillus subtilis 168 and W23 chromosomes	
PUBMED		Gene 98 (1), 107-112 (1991)	
COMMENT		Original source text: B.subtilis (strain W23) DNA. Draft entry and computer-readable sequence for [2] kindly submitted by R.G.Wake, 24-JUL-1990.	
FEATURES		Author address: R.G.Wake University of Sydney, Dept. of Biochemistry NSW 2006, AUSTRALIA E-mail: GERRY@EXTRO.UCC.SU.OZ.AU. Location/Qualifiers	
source		1. .3187	
		/organism="Bacillus subtilis"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:1423"	
source		1. .1499	
		/organism="Bacillus subtilis"	
		/mol_type="genomic DNA"	
		/strain="W23"	
		/db_xref="taxon:1423"	
		complement(120..1337)	
		/note="ORF 405"	
		/codon_start=1	
		/transl_table=11	
		/protein_id="AAA22720.1"	
		/db_xref="GI:143478"	
CDS		/translation="MTNQTSRKKRYANIIPMEELHSEKRDLPFFPIYDKLRRES VRYPDRCDVDFYDDVQVFLKNPLFSKRGIGTESILTMDFPKTKLRALVSRF TPKAVKLETRIKDVTAFLOAROKSTIDIEDFAGPLPVLIIAEMLGAPIEDRLHI KTYSDVLVAGKSDSKADVHNRDGHAFLSDFRDIILSKRABPKEDIMTMLIQ AEIDGELYTRSQLGFCILLVAGNETTNLIANVRILTSDSVVQQVQRQNTDVAN VIEETLRYSPVQAGRVATDETELGGVFIKKGSVLSWIASNRDEKFCXPDFCKI	

repeat_region	DRPSYPLHSFGFIHFC ¹ LGAPLARLEANIALSLLSMSACIEK ³⁴ RAHDEK ¹ EAIDPSPV FGVKRLPVRITFK ¹ 1461..1615 /note="inverted repeat" 1642..2010 /note="replication terminator protein (rtp; terC) " /codon_start=1 /transl_table=11 /protein_id="AAA22721.1" /db_xref="GI:143479" /translation="MKSEKPSGTGLVKQRAFL ¹ XYMTWTEQERLYGLKLLIEVLRSE FKETGKPNHTEVYRSLHLLDGLLQIKVKVKEGAKLQEVVLY ¹ QFADYEAALKYKQ LKVELDRCKKLEIKALSDNF" 2156..2929 /note="ORF 257 proC protein (put.); putative" /codon_start=1 /transl_table=11 /protein_id="AAA22722.1" /db_xref="GI:143480" /translation="MRTKTKRTKEMLPIDOKKVAFIAGSMAEGMISGIVRANKIPKQ NICVTRNTERLAELQYIKGASPNQICIDMDVLI ¹ LAMPKDAESALSSLKTRI OPHOLILSVLAGITTFIEQSLNQQPVVRVMPNTSMIGASATAIKGVKVEDLOK LAEALLGCMGEVTTIOENQMDIFGTIAGSPAFYVYLMEF ¹ EKTGEAGLDKQLSRSI GAQTLLGAARKMLMETGEQPEVLRDNTSPNGTTAAGLQA ¹ "	
CDS	Query Match Besc Local Similarity 53.8%; Pred. No. 4.9e-55; Matches 474; Conservative 0; Mismatches 407; Indels 0; Gaps 0;	
ORIGIN		
QY	241	ATCATTAACATGACCCCGCAACATCAAAATTCGGTTCAGTCGTGAACAAAGCCGTTT 300
DB	1091	ATATTAACAATGACCTCCGAAACACACCAATACGGGCTCTCGTCAGCAGACGGT 1032
QY	301	ACTCCGCGGTGATGAAGCAATGGAAACCGAGAAATTCAGAAATTCACAGATGAATG 360
DB	1031	ACCCAAAAGCAGTAAACAACTCGAAACCCGCATAAAGACGTGACGCGTTTCTCTTA 972
QY	361	CAAAAATTTACGGGCGGAGTGTTCACCTTGTTCAGATTTTCATACCGCTTCG 420
DB	971	CAAGAAGCAGCGCAAAAGAGACCAATCGATATCATTTGAAGATTTTCAGAGTCCCT 912
QY	421	GTATTGTGATCTGAGCTCTGGAGTGCCTTCAGCGCAGATGGAACAGCTTTAAAGCA 480
DB	911	GTATCATCATAGCTGAATGCTGGCGCTCGGATTTGAAGACCGCCACTTGATTAACA 852
QY	481	TGGTCTGATCTTCGTGCTACACCGAAGATTAAGTGAAGAGCTGAAAGACGCTTT 540
DB	851	TACTCCGACGTACTGTAGCTGGAGCAAGGACAGCTCCGACAAAGCTGTTGCTGACATG 792
QY	541	TTGGAAGAACGAGATAAGTGTGAGGAAGAACTGGCGCGTTTTCGCGCATCATAGAA 600
DB	791	GTCCATAATCGCGGTGACGCGCCACGCTTTTAAAGTGACTATTCAGAGACATCTGTCA 732
QY	601	GAAGAAGCAAAACAAACCGGAACAGGATATTATTTCTATTTAGTGGAAAGCGGAAGA 660
DB	731	AAACGAGAGCTGAACCAACAAAGAGATTAATGACAAATGCTTTTACAGCGCAAAATCGAT 672
QY	661	GCGGAGAAAGCTGTCCGTTGAAGAGCTGATTCGTTTTCACGCTGCTGCTGGTGGCGGGA 720
DB	671	GGAGAAATTTAAACAGAAAGACAGTGTATCGGTTTGTATTTCTCTCTCTCGCGAGGG 612
QY	721	AATGAAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTAGAAACGCGCAGGC 780
DB	611	AACGAAACGACTACAAATTTGATCGCAATGAGTTCGCTATCTGACAGAAAGATTCAGTG 552
QY	781	GTTTACGAGGAACTCCGACGCACTCTGAACTGATGCTTCAGCGAGTGGAGGAAGCCCTG 840
DB	551	GTCCACACGAGGTGAAGCAAAATACAGCAATGTTGCCAATGTAATGAAGAAACCGCTT 492
QY	841	CGTTTCAGAGCGCGGCCCGCGTTTTCAGSCGCATTCGCCAAGCGGATACCGAGATCGGG 900
DB	491	CGCTACTATTCGCTGTGAGGCGCATTTGGCGCTGTGCGACGGAAGACACAGAGCTTGGC 432

```
QY 901 GGGCACCTGATTAAAGAGTGATATGTTTGGCGTTTGTGCGATCGCGAAATCGTGAT 960
Dh 431 GGAGTGTTTCATCAAAAGAGGATCTTCOGTTATCAGCTGGATCGGTCAGCCATCGGAT 372
QY 961 GAAGCAAGTTTGACAGACCGACATGTTTGATATCGCGCCATCCCAATCGCATATT 1020
Dh 371 GAAGTAATTCGTAAACCTGATGTTTAAAGATTGACCGCCCTCTCTCATCTC 312
QY 1021 GCGTTTGGCCACGCGATCCATTTTCCCTGGGGCCCGCTTCCCGCTTGGAGCAAT 1080
Dh 311 AGCTTGGATTCCGGATTCACTTTTCCCTCGCGCTCCATTGCGCAGGCTAGAAGCAAC 252
QY 1081 ATCGGTTAAAGCTTGTGATTTCTGCTTTCCTCATATGA 1121
Dh 251 ATCGCTCTTTCCCTCCCTTTATCAATGTCAGCTGTGATCGA 211

RESULT 7
LOCUS BMCTP450A 4317 bp DNA linear BCT 13-OCT-1993
DEFINITION B.megaterium cytochrome P450meg, ORF1 and ORF2 genes.
ACCESSION Z21972
VERSION Z21972.1 GI:288298
KEYWORDS cytochrome P450meg; ORF1; ORF2.
SOURCE Bacillus megaterium
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1701 to 3400)
AUTHORS Rauschenbach,R., Isernhagen,M., Noeske-Jungblut,C., Boidol,W. and
Siewert,G.
TITLE Cloning sequencing and expression of the gene for cytochrome
P450meg, the steroid-15 beta-monooxygenase from Bacillus megaterium
ATCC 13368
JOURNAL Mol. Gen. Genet. 241 (1-2), 170-176 (1993)
MEDLINE 94049677
PUBMED 8232201
REFERENCE 2 (bases 1 to 4317)
AUTHORS Siewert,G.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-1993) Siewert G., Schering AG, Muellerstr.
170-178, D-1000 Berlin 65, Germany
FEATURES
source 1..4317
/organism="Bacillus megaterium"
/mol_type="genomic DNA"
/strain="ATCC 13368"
/db_xref="taxon:1404"
37..717
/gene="ORF1"
37..717
/gene="ORF1"
/citation=[1]
/codon_start=1
/transl_table=11
/db_xref="CAA9984.1"
/db_xref="GI:288298"
/db_xref="SPTREMBL:Q06073"
/translation="MRISLKKALQSLGGKXGTVTVSLMLLLFLNLVPLIVEVIGS
GGSEWLMQEBTPWSDIFSMVSIAPLITSTWFLNLVREGNPGIPEVFAIYKD
KTSFKLIGASILQAIIFLMSVLLIIPGIKAIAYSQQFFLLKDPHTVLEATES
KRMKGLAKVFLMHLISFIGWILGILGLLWLIIPVAGTTTAAFYBELLVQEDI
DDQQLIEG"
724..748
/citation=[1]
/rpt_type=INVERTED
/rpt_unit="724..733"
751..794
/citation=[1]
/rpt_type=INVERTED
/rpt_unit="751..771"
1330..1373
/citation=[1]

repeat_region
15.3%; Score 182.6; DB 1; Length 4317;
Best Local Similarity 51.1%; Pred. No. 2.7e-41;
Matches 567; Conservative 0; Mismatches 494; Indels 48; Gaps 4;

QY 52 AAAACAACACAGGATCGGTATCATCCGTTTCCATGATGATCGATGAGAGGATCGG 111
Dh 2063 AAAACAAGACGAGGAAATTTAGCCCGTACCGTTGTTAAAGGATGTTAGAAATGAC 2122
QY 112 CCGTTTCTCTTGATGACAAACCAAGTGTGAGCGTTTCTTTATGATGATGTCMAA 171
Dh 2123 CCGTGAGTTATCACAGGAAACGATACGTTGAAATGTCCTTAATATGAAGATGGAAG 2182
QY 172 AAGATTGTGGGATAAAGAGTTGTTTTCAGTTCCATGCCGACGACAGCAAGCTTATT 231
Dh 2183 CGGGTTCTCAGTGATTAACATTTTTCAGTTTTCGGAACGACGACGATTTTCAGTT 2242
QY 232 GGAAATTC-----ATCATTAACATGGACCGC 258
Dh 2243 GGAACGATAGTACGAGGATTTCTGTGCTGAAAAGATCCAAATCACTGAATCGGATCCA 2302
QY 259 CCGAGCATACAAAATCGTTTCAGTCGTAACACAGCCCTTACTCCGGCGCGTGATGAAG 318
Dh 2303 CCGTATCATAGAAACGCGTTTCACTGCTGGCAGCAGCATTCACACTAGAGTCTTCAA 2362
QY 319 CAATGGGACCGGAGATTCAAGAAATACAGATGAATCAAAATTTCAAAATTTAGGGCGCG 378

repeat_region
/rpt_type=INVERTED
/rpt_unit="1330..1347"
/citation=[1]
/rpt_type=TANDEM
/rpt_unit="1976..1988"
2008..2013
/citation=[1]
/function="steroid-15beta-monooxygenase"
/citation=[1]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="cytochrome P450meg"
/protein_id="CAA9984.1"
/db_xref="GI:288300"
/db_xref="GGA:Q06069"
/db_xref="SWISS-PROT:Q06069"
/translation="MKEVIAVKEITRFKTRTEBFSYAMCKRMLNDPVSHEGTDTW
NFKYBDVKRVLSDYKHFSVRKRTTISVGTDSBEGSVPEKIQTESDPDHRKRRL
LAAATPRLQNWEPRIOETADELIGQMDGTEDIVASLASPLPIIVMADLMGVPSK
DRLLFKKVDITLFPDREQBEVDKLAQAAYQYLYPIVYVQKELNADDIISDL
LKSEVDGEMFTDDEVVVTMLIIGAGVETSHLLANSFYSLDXDKKVPYDELHNLDL
VPOAVBEMLRFRFNLIKLDRTVKEDNDLLGVLEKSGSVVYVWMAAANDDEMFEDPT
LNTHRPNNKHLTFGNGPHEFCIGAPLARLEAKIALTAFLAKKFKHIEAVSFQLEENLT
DSATGQTLTSLPLKASRM"
3256..3303
/citation=[1]
/functions="terminator"
/rpt_type=INVERTED
/rpt_unit="3256..3275"
3560..4210
/gene="ORF2"
3560..4210
/gene="ORF2"
/citation=[1]
/codon_start=1
/transl_table=11
/protein_id="CAA9986.1"
/db_xref="GI:288301"
/db_xref="SPTREMBL:Q06074"
/translation="MKLQQLTIATIIISVLISLIGFSMAFTISANEYLKFDVIVSL
QWESPLLTDMKEFTYIGSPASIIISLVILFYLRLKHLRLVLFATVWVGSSELL
NLWKLFFQARPDHLRLIDIGYSPSGHAMNAFSLYGLTFLMRHITARWARILL
ILFSMLMLLSIGIRIYLVGHPYSDIILAGYLACGCMIAISLWFFQRYQDRKXKDR"
```

Db	2363	AACTGGGAACCTCGCATTCAGGAATTCAGATGAATGATGGACAAATGATGGTGA	2422
Qy	379	AGTGAATTTGACCTTGTTTACGATTTTTCATACCGCTTCCGGTTATTTGATATCTGAG	438
Db	2423	ACGGAAATCGATATTTGGCATATGGGAGTCCGCTCCGATCATTTGTATGCGCGAT	2482
Qy	439	CTGCTGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAGCATGCTGTGATCTTCGTGC	498
Db	2483	TGATGGGGTCCCTCGAAATCGTTTATTGTTTAAAGAAATGGTGGATACCTTAITTT	2542
Qy	499	AGTACACGAGGATAAAGTGAAGAGCTGAAAAGCCTTTTGGAGAACGAGATAG	558
Db	2543	CTTCCTTTTATAGAGAAAGCAAGAGAAAGTAGATAATTGAAGCAAGTTCAGCA--	2600
Qy	559	TGTGAGGAGAACTGGCGCGTTTTCGCCGATCATAGAGAAAGCGAAACCAACCG	618
Db	2601	-----AAGAAATCTATCAGTATTTGATTCGATTTTGTGTCNAAACGATTGAACCCG	2653
Qy	619	GAACAGGATATTTCTATTTAGTGAAGCGGAAGAAACAGGCGAGAGCTGTCCGTT	678
Db	2654	GGGATGATATCATCTCATCTATTGAAGTCCGAAGTGGATGGGGAATGTTTACGGAT	2713
Qy	679	GAAGAGTGAATTCGGTTTGGAGCTGCTGCTGCTGCGCGGAATGAACCACTACAAAC	738
Db	2714	GATGAGTTTGTCCGACGACCATGCTGATTTTAGTGGAGGAGTCGAGACAACCATCAT	2773
Qy	739	CTGATTTCAAATTCGATGTACA---GCATATTAGAAACCGCAGCGCTTTACGAGGAAC	795
Db	2774	TTATGGCCATAGCTTTTATTCGCTGCTATATGATGACAAAGAGTTTATCAAGAGTTA	2833
Qy	796	CGCAGCATCTCTGAATGATGCTCAGGAGTGGAGGAGCCTTGGTTTCAGAGCGCG	855
Db	2834	CATGAACACTGGATTAGTTAGTTCGCGAGCGGTGGAAGAAATGCTCCGTTTCCGATTCAT	2893
Qy	856	GCCCGGTTTGGAGCGCATTCGACGCGGATACGAGATCGGGGCGACCTGATTA	915
Db	2894	CTTATTAATTTGGATCGCATGTAAAGGAAGATAACGATCTATGGAGTGGAAATGAAA	2953
Qy	916	GAAGTGATATGTTTGGCGTTTGGCATCGGCAATCGTATGAAGCAAAAGTTTGAC	975
Db	2954	GAAGGGATAGCTGTTTGGATGAGTGCAGCTAATATGACGAGAGATGTTGAA	3013
Qy	976	AGACCCACATGTTTGATATCCCGCGCATCCCAATCCG---CATATGCGTTTGGCCAC	1032
Db	3014	GACCCCTTCACACTTAATATCCACCGCCCTAATAATAAGAAACATCTCACATTCGGTAAT	3073
Qy	1033	GGATCCATTTTGCCTTGGGCGCCGCTTGGCGCTTGAAGCAATATCGCGTTAACG	1092
Db	3074	GGCCCTCATTTCTCCCTCGGAGCACCGCTAGCAGGCTGGAAGGAAGATTCGCTTACT	3133
Qy	1093	TCCTTGATTTCTGCTTTTCTCATATGGA	1121
Db	3134	GCATTCCTGAAGAAATCAAGCATTTGA	3162
RESULT 8			
BMC4501			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
BMC4501			
1902 bp DNA linear BCT 08-AUG-1995			
Bacillus megaterium gene for cytochrome P-450 (BM-1).			
X16610			
GI:39626			
cytochrome; cytochrom			
Bacillus megaterium			
Bacillus megaterium			
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
1 (bases 1 to 1902)			
He, J.S., Ruettinger, R.T., Liu, H.M. and Fulco, A.J.			
Molecular cloning, coding nucleotides and the deduced amino acid			
sequence of P-450BM-1 from Bacillus megaterium			
Biochim. Biophys. Acta 1009 (3), 301-303 (1989)			
90089408			
2597681			

REFERENCE	2	(bases 1 to 1902)
AUTHORS	Liang, O. and Fulco, A.J.	
TITLE	Transcriptional regulation of the genes encoding cytochromes	
JOURNAL	P450BM-1 and P450BM-3 in Bacillus megaterium by the binding of	
MEDLINE	Bm3R1 repressor to Harbie box elements and operator sites	
PUBMED	J. Biol. Chem. 270 (31), 18606-18614 (1995)	
REFERENCE	7629191	
AUTHORS	3 (bases 1 to 1902)	
TITLE	Fulco, A.J.	
JOURNAL	Direct Submission	
FEATURES	Submitted (27-AUG-1989) Fulco A.J., University of California, Lab.	
source	Biomedical & Environmental Science, 900 Veteran Ave, Los Angeles CA	
	90024, U S A	
	Location/Qualifiers	
	1..1902	
	/organism="Bacillus megaterium"	
	/mol_type="genomic DNA"	
	/strain="ATCC 14581, isolate Fulco PB85"	
	/db_xref="taxon:1404"	
	/clone="BM1-1.9"	
	/cell_line="E.coli DH5alpha"	
	/clone_lib="pUC19"	
	322..327	
	/note="35 region"	
	346..398	
	/note="inverted repeat"	
	346..351	
	/note="10 region"	
	489..493	
	/note="ribosome binding site"	
	505..1737	
	/note="unnamed protein product; cytochrome P450 (BM-1) (AA	
	1-410)"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="CAA34612.1"	
	/db_xref="GI:39627"	
	/db_xref="GOA:P14762"	
	/db_xref="SWISS-PROT:P14762"	
	/translation="MNKVIPTVETPKFQSRABEFFPIQYKEMLNKSPVYFHESTNT	
	WNYFOYEHVQVLSNDYDFSSDGRRTTFVGDNSKKKSTSEITNLNLDPPDHRKRS	
	LIAAFTPRSLKNWEPRIKQIADLVEAIKNSTINIVDLSPPSLVADLPFGVPV	
	KDRYQKXWDILPQYDORLEIEOKORAGAEYFQYLYPIVIERKSNLSDIISD	
	LIOAVDGEFTBEIIVHATMLLGACVETSHAIANMFYSLVDKSLYSELNRR	
	LAPAVDEMRYKHFHSRRRTYKQDNELLGVKLKGDVVIAMWSACNMDSFENPF	
	SVDIRPTNKKHLTFGPHPCFGLGAPLARLEMKILAEFLAEFSEHIEFEFELEPHL	
	TASATGQSLTYLPMTVYR"	
ORIGIN		
Query Match	14.2%;	Score 169; DB 1; Length 1902;
Best Local Similarity	50.5%;	Pred. No. 2.1e-37;
Matches	554; Conservative	0; Mismatches 495; Indels 48; Gaps 4;
Qy	64	GATCGGTATCATCCGTTTCCATGGTATGAATCGATGAGAAAGGATGCGCCCTGTTTCTTT 123
Db	555	GAAGATTTTCCCTATCCATGTTTCAACAGAAATGTTGAACAACTCCCTCTGTTTATTTT 618
Qy	124	GATGAAGAAACCAAGTGTGGAGGTTTCTTTATGATGATGTCACAAAAGATTTGTCGG 183
Db	619	CATGAAGAAACCAACACATGGAATGTTATCCAAATCGAACATGTTTAAACAAGTCTTGAGC 678
Qy	184	GATAAGAGTGTGTTTCCAGTTGATCGCGCAGCAGCAAGCTCTATTGGAATTCCTCA-- 241
Db	679	AACTACGATTTTTTTCAAGCATGGACAGAACCATCTATTTTGTGAGATATAGT 738
Qy	242	-----TCATTAACATGGACCCCGCAGCATACA 270
Db	739	ANGRAAAAAGACATCTCCATCACCACCACTTACAAACTTAGATCTCTCGACCAACCG 798
Qy	271	AAATCCGTTTCACTGCTGTAACAAAGCCTTTACTCGCGGTGATGAACCAATGGGACG 330
Db	799	AAAGCGCATCTTTCCTTGTCTGACCCCTTACTCTCTCGAGTTTAAAGAACTGGGAACG 858

331 AGAATTCAGAAATCACAGATGAACCTGATTCAAAATTTTCAGGCGCGAGTCAGTTGAC 390
|||||
859 CGAATCAACAATTCGAGCAGACCTTGTAGAGCGATACAAAAGAAATCAACTAAT 918
|||||
391 CTGTTTCACGATTTTTCATACCGGCTTCGGGTATTTGTGATATCTGAGCTGCTGGAGTG 450
|||||
919 ATTGTGGACGATTTATCT 978
|||||
451 CTTTCAGCGCAGATGACGATTTTAAAGATGCTGATCTCTCTCTCTCTCTCTCTCTCTCTCT 510
|||||
979 CCGGTAAAGACCGATATCAGTTTAAAGATGCTGATCTCTCTCTCTCTCTCTCTCTCTCTCT 1029
|||||
511 GATAAAGATGAGAGAGCTGAAAGACCTTTTGGAGAAACGAGATPAAGTGTGAGAGAA 570
|||||
1030 CTTATGACCAAGAGATTTAGAGAGATTTGAACAGAAAGCAACGTCGAGAGCTGAA 1089
|||||
571 CTGGCCGCTTTTTCGCGGATCTAGAGAAAGCGGAAACAAACCGGACGAGATATT 630
|||||
1090 TATTTTCAATACCTCTACCCGATTTGCTATTGAAAGCGGCTCCAACTTTCTGATGATATT 1149
|||||
531 ATTTCTATTTAGTGAAGCGGAGAAACAGCGGAGAGCTGTCGCGTGAAGAGCTGATT 690
|||||
1150 ATCTCAGACTTAATTCAGCTGAACTTGAATGATGATGAACTTACAGATGAAGAAATTGTA 1209
|||||
691 CCGTTTTCACGCT 750
|||||
1210 CACGCTACCATGTTGCTCTAGTGCAGCGCTTGAACCCACAACTCAAGCTCAAGCTCAAT 1269
|||||
751 CGATGATGACGATTTA---GAAACGCGCGCTTTACGAGGAGCTCGCGAGCCATCTCT 807
|||||
1270 ATGTTTATTTCTTTTATATGACCAAGCTGTTATACGCGAGCTGAGAAACATAGA 1329
|||||
808 GAATGATGCTCAGCGAGTGAAGAGCTTTCGCTTTTCAGAGCGCGCCCGCTTTTG 867
|||||
1330 GAATGATGCTCAGAGAGTGAAGAGCTTTCGCTTTTCAGATGCTTCAATCTCCAGAGCA 1389
|||||
868 AGGCGCATGCGAAGCGGATACGAGATCGGGGCGGACCTGATTAAGAGAGTGAATG 927
|||||
1390 GATCGACCGTTAAACAGAAACGAGCTATTAGCGGTTTAACTAAAGAGAGATGTC 1449
|||||
928 GTTTGGGCTTTGGCATCGCAATCTGATGAGCAAGATTTGACAGCGCGACATG 987
|||||
1450 GTTATGCGTGATGAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1509
|||||
988 TTTGATATCCGCGCATCCCAAT---CGCATATTTGGCTTTGGCGCACCGCATTCATTTT 1044
|||||
1510 GTAGACATTCATGCTCCAACTTAATAAAGACCTTAACATTCGAGAAACGACCTCATTTT 1569
|||||
1045 TGCCTTGGGCGGCGGCTTGGCGGCTTGAAGCAATATGCGGTTAAAGCTTTGATTTCT 1104
|||||
1570 TGTTTGGGCGGCGGCTTGGCGGCTTGAAGCAATATGCGGTTAAAGCTTTGATTTAGAG 1629
|||||
1105 GCTTTTCTCATATGGA 1121
|||||
1630 GCTTTTCTCATATGGA 1646
|||||

RESULT 9
AB072568
LOCUS
DEFINITION
Streptomyces virginiae visa, visB, visC, visD genes for L-lysine
2-aminotransferase, 3-hydroxypicolinic acid:AMP ligase, lysine
cyclodeaminase, cytochrome P450 monooxygenase, complete cds.
AB072568.1 GI:18150120
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Streptomyces virginiae
Streptomyces virginiae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1
AUTHORS
Namwat,W., Kamioka,Y., Kinoshita,H., Yamada,Y. and Nihira,T.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL
TITLE
source
FEATURES
1..6629
/organism="Streptomyces virginiae"
/mol_type="genomic DNA"
/db_xref="taxon:1961"
complement(611..1870)
/gene="visA"
complement(611..1870)
/gene="visA"
/note="visA"
/codon_start=1
/transl_table=11
/product="L-lysine 2-aminotransferase"
/protein_id="BAB83671.1"
/db_xref="GI:18150121"
/translation="MTAAQRYLVPPFRRGSIQVQADLAALASVHSAEPLSGWNR
FEQPREVYTGARHLSVSGTVALEVLALDLPQGVDEVIATPQTQATVQLLDH
TVFCDVDPDTLMDPQVLDLITDTRKAILLVHYGGIPADMAAILARARRAGVWVE
DCAHGASGSGALGDIQSCFHSHTKNTTLGEGMITLNADNRVERLRGN
ETGVGFAPVGFADDEPALLPMWRPSADYRQAVGISGAGTINATMSEAAAGVQL
DALERTRRRIARLDAASAFPRHPPHAGSGSTYHLYTFETVAGRGASDSC
APLDRGVEVQLRYFLLHLEVAARPRSGRVPQARVWFDEHLNLPCHPLGTDEQV
DHLVESVTAAMHEATARTCEPLAVR"
complement(1867..3522)
/gene="visB"
complement(1867..3522)
/gene="visB"
/note="visB"
/codon_start=1
/transl_table=11
/product="3-hydroxypicolinic acid:AMP ligase"
/protein_id="BAB83672.1"
/db_xref="GI:18150122"
/translation="MLDGCVPMPDPAAREYRAGYWRGETLGLLRGWAERYGDRVAL
VGDDGTRVYRELDMMCDLAAAGPAGQVAPRERVVLQPNTPSPFVAVCFALRUGAV
PVFALPSYAVRSCGTRSCPVRSRWIPEDEVRGFDRALVRLJELPRVAVRVFVAGD
PGEDHGVALLAALQEDPVLPDPAIDVAFLLSGGTTALPKM:PRTHDDYAYQTRI
TAGICELGEDTVYLAIVLVEFNPMGCPGIIQVLGVRVVPAPDPTPOTCFPLIERE
RYTMSVIFTIVHMLDAAGRTGHDLSLDVYVQSSKLHDEVAARIEPELGVRLQOV
FQMAEGLTFTTYDDPRQVLTQGRPVSPADEIRVWDGDDREVDGCTVGLLTGPGY
TLRGYRAPEHNASATADGFYRSGDVRRTPTGRIIVGVRVQDVGIRGDKVSAFV
EGHLYAHPEVQQAQAVKAPDRVLGKICAYVVAADRGAPTLFGVRRLLRAGLADF
LPDRLEVVEVPLTGLNKVKVLAARAGLLHSGGAGAGAR"
3759..4646
/gene="visC"
3759..4646
/gene="visC"
/note="visC"
/codon_start=1
/transl_table=11
/product="lysine cyclodeaminase"
/protein_id="BAB83673.1"
/db_xref="GI:18150123"
/translation="METWLSRDVTRAVESVGRJALMDRVIDRLTEGLAEVGRGAR
SPARGNREKVPVGIWEWMPHRESGDSITKAVGSPANPGRFGLTIGTVARFDD
TTGSLTALIDGVLTAIRTGASAFASRLAAPDSRVVGLIGSQSVTQLHALTRV
EVRVLTWDADPSHAASLADRVAFGLQVQAEPRHIAEADIICTATVPVGGPVL
PADAVRPHLHINAGRLVGLKTELEFLDLIRRAHVSPPHPEARHEGECQQLPEHIG
LSLSELCDAPREPPTLCAAD"
4994..6202
/gene="visD"

Characterization of virginiamycin S biosynthetic genes from
Streptomyces virginiae
Gene 286 (2), 283-290 (2002)
21940635
11943483
2 (bases 1 to 6629)
Namwat,W., Nihira,T. and Kinoshita,H.
Direct Submission
Submitted (05-OCT-2001) Wises Namwat, Graduate School of
Engineering, Osaka University, Biotechnology; Yamada-oka, Suita,
Osaka 565-0871, Japan [E-mail:noi4bio.eng.osaka-u.ac.jp,
Tel:81-6-6879-7433, Fax:81-6-6879-7432]
Location/Qualifiers

Qy	Db	Qy	Db	RESULT 10
1011	TCGCATATTGGCTTTGGCCACGGCATCCATTTTCCCTTGGGGCCCCGCTTCCCGTCT 1073	1071	TGAAGCAATATCGC 1085	BSCYPAZL
6007	CCCTCAGCTGGCTTCGGCAGGCATCCACTACTGCTTCGGCGGTTCGGTGGCGGCT 6066	6067	GGAGGCGAGGCTCGC 6081	LOCUS
				DEFINITION
				BSCYPAZL
				Bacillus subtilis cypA, azlB, azlC and azlD genes, ORF49, ORF91 and
				ORF182.
				ACCESSION
				Y11043.1
				VERSION
				Y11043.1
				KEYWORDS
				azlB gene; azlD gene; barstar-like protein; cypA gene;
				cytochrome P450-like protein; ORF182; ORF49; ORF91; transcriptional
				regulator.
				SOURCE
				ORGANISM
				Bacillus subtilis
				Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
				REFERENCE
				1
				Belitsky,B.R., Gustafsson,M.C., Sonenshein,A.L. and Von
				Wachenfeldt,C.
				An lrp-like gene of Bacillus subtilis involved in branched-chain
				amino acid transport
				J. Bacteriol. 179 (17), 5448-5457 (1997)
				97431495
				PUBMED
				9287000
				REFERENCE
				2 (bases 1 to 5008)
				von Wachenfeldt,C.
				Direct Submission
				Submitted (04-FEB-1997) C. von Wachenfeldt, Lund University,
				Department of Microbiology, Solvegatan 12, S-22362 Lund, SWEDEN
				revised by submitter 02-APR-97
				Sequence is continuous with U79494.
				FEATURES
				source
				1..5008
				/organism="Bacillus subtilis"
				/mol_type="genomic DNA"
				/strain="1A1"
				/db_xref="taxon:1423"
				/map="233 degrees"
				/clone_lib="pMG001"
				<1..551
				/note="ORF182"
				/codon_start=3
				/transl_table=11
				/product="hypothetical protein"
				/protein_id="CAA71935.1"
				/db_xref="GI:1926276"
				/db_xref="GOA:O07081"
				/db_xref="SPTREMBL:O07081"
				/translation="ALLIVDVOKAFDDKGMERNVKAENISKILELWREKGVTVII
				IQHTSDKPSHLSFHPKNGSFAIKIVKPMDEVIITKTWSSFIGTLEEFKLNKEIT
				WITGLTPHCVSTTTRMSGNLGFTYLLISATAAFGMQRQNDITYDAATIHNIISLA
				WITGLTPHCVSTTTRMSGNLGFTYLLISATAAFGMQRQNDITYDAATIHNIISLA
				LHDEFATILTDDQINDFIKTH"
				650..799
				/note="ORF49"
				/codon_start=1
				/transl_table=11
				/product="hypothetical protein"
				/protein_id="CAA71936.1"
				/db_xref="GI:1926277"
				/db_xref="SWISS-PROT:O0794E"
				/translation="MFLYRQARAEVSKVAKLFHQSFNQYPMDFLLVKREKKIHLF
				KCIKT"
				1192..2424
				/gene="cypA"
				1192..2424
				/gene="cypA"
				9287000
				LOCUS
				DEFINITION
				BSCYPAZL
				Bacillus subtilis cypA, azlB, azlC and azlD genes, ORF49, ORF91 and
				ORF182.
				ACCESSION
				Y11043.1
				VERSION
				Y11043.1
				KEYWORDS
				azlB gene; azlD gene; barstar-like protein; cypA gene;
				cytochrome P450-like protein; ORF182; ORF49; ORF91; transcriptional
				regulator.
				SOURCE
				ORGANISM
				Bacillus subtilis
				Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
				REFERENCE
				1
				Belitsky,B.R., Gustafsson,M.C., Sonenshein,A.L. and Von
				Wachenfeldt,C.
				An lrp-like gene of Bacillus subtilis involved in branched-chain
				amino acid transport
				J. Bacteriol. 179 (17), 5448-5457 (1997)
				97431495
				PUBMED
				9287000
				REFERENCE
				2 (bases 1 to 5008)
				von Wachenfeldt,C.
				Direct Submission
				Submitted (04-FEB-1997) C. von Wachenfeldt, Lund University,
				Department of Microbiology, Solvegatan 12, S-22362 Lund, SWEDEN

KVLPASAVIGLILVYCKVDVSLSSGHGIPELVCAAAVAVLLHLWKOMLLSIAGGTGVY
KVLVQLVF

Query Match	13.6%	Score 162.4	DB 1	Length 5008
Best Local Similarity	51.4%	Pred. No. 2.1e-35		
Matches 463	Conservative 0	Mismatches 416	Indels 21	Gaps 3
227	CTATTGGAAATTCATCATTAACATGAGCCGCGGAGCATACAAAATCGGTTCACTCG	286		
1463	CTTTAAACAAACATATGCTTAATTCAGATCCCTGATCAACACCGCTTCGAACTCTG	1522		
287	TGAACAAAGCCTTACTCCCGCGTGTAGCAATGGAACCGAGAAATCAAGAAATCA	346		
1523	TTCAAAAGCAATTTACGCATCGGATGATTTTGCAGTTTGGAGACAAATTCAGCATATCG	1582		
347	CAGATGAACCTGATTCAAAAATTTACGGGGCGGAGTGTGATTTGACCTTSTTCACGATTTT	406		
1583	CAGATCTTTTATGGATTAAGTGCAGCCCAACAAATTCATGACCTAGTAGAGGATTATG	1642		
407	CATACCCGCTTCCGGTTTATTTGTATATCTGAGCTGTGGAGTGCCTTCAGCGCAGATGG	466		
1643	CGTTTCCATTGCCGATTATCGTGAATGTAATGCTCGGCATTTCCCTTGGAGATCGTC	1702		
467	AACAGTTTAAAGCATGCTCTGATCTTCTGTGACGTACACCGAGGATAAAAGTCAAGAA	526		
1703	AAAAATTCAGATGATGCTGCCA-----AGCAATTAATGATTTTTCAGATGCCCC	1750		
527	CTGAAAAGCCTTTTGGAAAGACGAGATAAGTGTGAGGAAGAACTGCGCGGTTTTTGTG	586		
1751	CTGAAAGA-----CTTCAGGAGATGACCACTTTTGTAGGGAGCTTTTGTGATATCTGG	1804		
587	CCGGCATCATAGAAGAAAGCGAAACAAACCGGAACAGGATATATTTCTATTTTAGTGG	646		
1805	AATCCTTAGTTTCCCAAAAAGAGAGAGAACTCGCAGGAGATTTAATCAGTGCATTAATTC	1864		
647	AAGCGGAAGAACAGGCGGAGAGAGCTGTCGGTGAAGAGCTGACTCCGTTTTTGGCAGCTGC	706		
1865	AAGCGGAAGTGAAGGCGACACAAATTAAGTACAGAAAGATTTGCTACTCGATGATTATGCTGC	1924		
707	TGCTGTGGCGGAAATGAAACCACTACAAACCTGATTTCAAAATCGGATGACAGCATAT	766		
1925	TTATCGTTGAGGTCATGAACACACCGTAACTTAATCAACCAATATGACATATGCTCTGA	1984		
767	TAGAAACCCAGGCGTTTACGAGGAACTGCGGAGCCATCTGAACTGATGCTCAGCAG	826		
1985	TGTGTACCATGATCACTAGAAAAGCTCCGCGACGAGCTTGATGAATTCAGGCA	2044		
827	TGGAGGAAGCCTTGGCTTTCAGAGCGCGCGCGCTTTTGGAGGCGCATTCGCCAGCGGG	886		
2045	TAGAGGAAGCATTCGGTTTTCAGAGCCTTGTAGCTAACGCAATTCGCTGGAAGGCTG	2104		
887	ATAGCGAGATCGGG---GGGCACTGATTAAGAAGGTGATGTTTGGCGTTTGTGG	943		
2105	AACCGTTTATCTGATCGCGCAGGAGATTAACGGAAGATGTGATCATCATTTCTTTAG	2164		
944	CATCGGCAAAATCGTATGAGCAAAAGTTTGAACAGCCGACATGTTGATATCCGCCCC	1003		
2165	CATCAGCAACCGCTGATGAAAAGATTTTCCAAATTCGCGACATATTTGATATTTGAAGAA	2224		
1004	ATCCCAATTCGCGATATTTGGTTTGGCCAGCGCATCCATTTTTCCTTGGGGCCCGCTTG	1063		
2225	AGAACACCGGCATATGCTTTTGGCCCATGGCAATCATTTTGTCTTGGTGCACAGTAG	2284		
1064	CCCGTCTTGAAGCAATATCGGTTTAAAGTCTTTTGTGATTTCTGCTTTTCTCATATGAGT	1123		
2285	CCCGCTTAGAGCAAGATGCGATTTCCACCTTATAGCGCGTGTCCCAATATACAGT	2344		

RESULT 11
BSU93876
LOCUS
DEFINITION Bacillus subtilis aminoglycoside 6-adenylyltransferase (aadK) gene,
BSU93876 17828 bp DNA linear BCT 12-OCT-1997

```

/codon_start=1
/transl_table=11
/product="cytochrome P450-like enzyme"
/protein_id="CAA71937.1"
/db_xref="GI:1926278"
/db_xref="GOA:008469"
/db_xref="SWISS-PROT:008469"
/translation="MSSEKKSVTILTESQLSSAFKDYEFYKELRSQALYPLSL
GALGKGLLSRVDADHLKLNKELKKYENVSPTAKEERPAALLKNETLTKHMLNSDPP
NLRRLTVOKAFDTHMILOLEDKIOHIAEDSLDKQVKNFPMNVDDYAFPLIIVIS
EMLGIPLDEQRQFVWSQAIIFSDAPERLQENDHLLGFEVLESILVRKQREPADG
LTSALQASSEGQSLSTELYSIMILLI VAGHETTVNLITNTMYALMCHDDOLEKURQ
QPLMNSATEBALRPHSPVELTTRWTAPFPIIIGQEIIRKQVITISLASANDEKIF
PRADIPDIRKKNRHAFGHNHFCILGAQLARLEAXIAISTLLRCPNIQLKGEKQM
KWKGNFLMRALDELPISF"
2696..2971
/notes="ORF91"
/codon_start=1
/transl_table=11
/product="hypothetical barstar-like protein"
/protein_id="CAA71938.1"
/db_xref="GI:1926279"
/db_xref="GOA:007938"
/db_xref="SPTREMBL:007938"
/translation="MRKIIIDGRDFENIEVLHDDKDKLIDFPDYGRNLDAWDCLTG
WDPLPLVLANKFNSFTLGSYADDVLEVIQEAQELKDFKIIIE"
3305..3778
/gene="azlB"
3305..3778
/gene="azlB"
/functions="transcriptional regulator of the Lrp/AsnC
family; repressor of the azlBCDEF operon"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="transcriptional regulator"
/protein_id="CAA71939.1"
/db_xref="GI:1926280"
/db_xref="GOA:007920"
/db_xref="SWISS-PROT:007920"
/translation="MGIMLDETDAKILRDIQEDASINLNLKSKIGLSPSACLARTQN
LVGSHDIKKTITVDEKGLGLEVTALALINLPLNRETTHSFLEIDINKFPQVQCYTL
TGSHDYMLKIVAKDMESYRNFIDSLMNPANISGVDTSIVNSTEKTQVSPIDEM"
3791..4555
/gene="azlC"
3791..4555
/gene="azlC"
/notes="product similar to Haemophilus influenzae
hypothetical protein; HII1738"
/codon_start=1
/transl_table=11
/protein_id="CAA71940.1"
/db_xref="GI:1926281"
/db_xref="GOA:007942"
/db_xref="SWISS-PROT:007942"
/translation="WNNKBSLSLSSPAINTHMKRSQIWAARSAFPTIPIFAGEL
FLGTAYGIPHSLSGSAIYPIIMSGMIFAGMEFVAANFLGAFNPMNALPLTLMVNA
RHLFGYSMLDKYGTGKKLYLFGCDSEFSINYTANVPANVDKGMFEVTLINH
LYVWAGAAIGGIFGYSYKFNTEGLDFVNTALFIVIFIQWMEKKHYSALTGLGLSVA
SLILFGNQOFTIPAMLAIIIGLVTLVRKPLEKAESV"
4552..4884
/gene="azlD"
4552..4884
/gene="azlD"
/notes="product similar to Haemophilus influenzae
hypothetical protein; HII1737"
/codon_start=1
/transl_table=11
/protein_id="CAA71941.1"
/db_xref="GI:1926282"
/db_xref="GOA:007923"
/db_xref="SWISS-PROT:007923"
/translation="MTMTMQMITIAMVVLGTMLTRFLPFMFPSGKPTPKYQVLG

```

partial cds, and YrdA (yrdA), YrdB (yrdB), hypothetical protein YrdC (yrdC), YrdD (yrdD), hypothetical cytochrome P450 protein YrdE (yrdE), ribonuclease inhibitor (yrdF), regulatory protein YrdG (yrdG), hypothetical protein YrdH (yrdH), hypothetical protein YrdI (yrdI), amino acid transporter (yrdJ), YrdK (yrdK), LysR family regulatory protein YrdL (yrdL), YrdN (yrdN), cation transport protein YrdO (yrdO), hypothetical protein YrdP (yrdP), lysR family transcription regulator YrdQ (yrdQ), hypothetical protein YrdR (yrdR) and hypothetical protein YrdK (yrdK) genes, complete cds.

U33876
 VERSION U33876.1 GI:1934641
 KEYWORDS
 SOURCE
 ORGANISM Bacillus subtilis
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (bases 1 to 17828)
 AUTHORS Sorokin, A., Bolotin, A., Purnelle, B., Hilbert, H., Lauber, J., Dusterhoft, A. and Ehrlich, S.D.
 TITLE Sequence of the *Bacillus subtilis* genome region in the vicinity of the *lev* operon reveals two new extracytoplasmic function RNA polymerase sigma factors SigV and SigZ
 JOURNAL Microbiology 143 (Pt 9), 2939-2943 (1997)
 MEDLINE 97453473
 PUBMED 9308178
 REFERENCE 2 (bases 1 to 17828)
 AUTHORS Hilbert, H., Lauber, J., Bolotin, A., Sorokin, A., Dusterhoft, A. and Ehrlich, S.D.
 TITLE Direct Submission
 JOURNAL Submitted (12-04-1997) Genetique Microbienne, CRJ INRA, Domaine de Vilvert, Jouy en Josas 9352, France
 FEATURES
 source
 1. .17828
 /organism="Bacillus subtilis"
 /mol_type="genomic DNA"
 /strain="1168"
 /db_xref="taxon:1423"
 /map="235 degrees"
 1. .787
 /gene="aadK"
 /gene="aadK"
 /note="similar to S.aureus plasmid pS194 Str protein"
 /codon_start=2
 /transl_table=11
 /product="aminoglycoside 6-adenylyltransferase"
 /protein_id="AAB80893.1"
 /db_xref="GI:1934642"
 /translation="VTLEGRSTWRNIPDPNDYDISYFVTVDSFKENDQWLEIFCK RIMQKPEDMELPPELGNWFSYIILEPDKGLDLTLPIREADYFANNGLVKVLL DKDSFINYKTPNDQYWKRPAREFDDCCNEFWVSTTVVGLARNEILFALDHNL EIVRNLRRMAWHASQKGYFSMGKNYKPKRYLSNKWEELMSTYSVNGYQSMWK SLFTCYALFRKYSKAVSEGLAYKYPDYDEGITRYTEGIYCSVK"
 1013. .1516
 /gene="yrdA"
 1013. .1516
 /gene="yrdA"
 /codon_start=1
 /transl_table=11
 /product="YrdA"
 /protein_id="AAB80894.1"
 /db_xref="GI:1934643"
 /translation="MFQTLDFLKSWEFEADATQKLNLSLTDSELSKQETISQNWTLGR IAWHTVAAIGITISNTDLTPQAPAEYFVPTSAFIADSYHQASNAFVQALKTWTH TQGRINFINIQOMPNGSLMLFELIQHNNHRRGQTVLMRQAGLTVPYGIYGAKEWAKF GLRAPKM"
 2326. .2697
 /gene="yrdB"
 2326. .2697
 /gene="yrdB"
 /codon_start=1
 /transl_table=11
 /product="YrdB"
 /protein_id="AAB80895.1"
 /translation="MFQTLDFLKSWEFEADATQKLNLSLTDSELSKQETISQNWTLGR IAWHTVAAIGITISNTDLTPQAPAEYFVPTSAFIADSYHQASNAFVQALKTWTH TQGRINFINIQOMPNGSLMLFELIQHNNHRRGQTVLMRQAGLTVPYGIYGAKEWAKF GLRAPKM"
 2326. .2697
 /gene="yrdB"
 2326. .2697
 /gene="yrdB"
 /codon_start=1
 /transl_table=11
 /product="YrdB"
 /protein_id="AAB80895.1"

/db_xref="GI:1934644"
 /translation="MEKLNQTNLLIRFTLEIAALISLGVYAMISPNCEYKXVLTLP
 IAWHVSFAVPHDPSPSGQTVIANGVTRLVIELLIFAMAVAAULFSYIKPYSIYF
 LCIHIIHYSIABRIKWLNLQ"
 2709. .2728
 /note="transcription terminator"
 2926. .3489
 /gene="yrdC"
 2926. .3489
 /gene="yrdC"
 /note="similar to *Arthrobacter* sp. N-carbamoylsarcosine
 amidohydrolase"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein YrdC"
 /protein_id="AAB80896.1"
 /db_xref="GI:1934645"
 /translation="MSEKALIIIVDQKADKKWGNVNVKAEENISKILEWRKRG
 WTVIYQHTSDKPHSLPHKNGEFAIKVPMDEVIITKNVSSFTGNLEFLKL
 NEITVITGLITPHCVSTTRMSGNLGFDYILSDATAAFGMEDQNDTYDAATIH
 ISATLHDEFATILITDQLNDFIKH"
 3588. .3737
 /gene="yrdD"
 3588. .3737
 /gene="yrdD"
 /codon_start=1
 /transl_table=11
 /product="YrdD"
 /protein_id="AAB80897.1"
 /db_xref="GI:1934646"
 /translation="MFYQCARADEVSKVAKLHQSFNQYPFMDLLVREKKIHLFL
 KCIKY"
 4130. .5362
 /gene="yrdE"
 4130. .5362
 /gene="yrdE"
 /note="similar to *Saccharopolyspora erythraea* cytochrome
 P450CvIIb1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical cytochrome P450 protein YrdE"
 /protein_id="AAB80898.1"
 /db_xref="GI:1934647"
 /translation="MSSKEKSVTILTESQLSSPAFKDEAYEFYKELRKSQALYELSL
 GAIGKGLISRYDDAIHLKNEKKKNYFVFAKXRPALLKNEETLTKMLNSDPP
 DHNRLTVQKAFTHRMVLOLEDKIQIADSLLEKQPNKFMNIVDDTAFPLPVIS
 EMGLPIEDQKRWQSAIIFSDAPERQENDHLLGSFVEYLESVVRKRRPAGD
 LISALIQASEGTOLSTELYLMILLIVAGHETTVNLTNTMTVCHHDQLEKLQ
 QPOLMNSAIEEARFSPVELTTIRWTAEPFILHQEIKRKDVIIISLASANREKIF
 PNADIPDIERKNNRHIAFGHGHFCLGQALRLAKIAISTLLRRCNPVQLKGEKKQ
 KWGNFLMRALDELPISF"
 5634. .5909
 /gene="yrdF"
 5634. .5909
 /gene="yrdF"
 /note="similar to *Bacillus amyloliquefaciens* ribonuclease
 inhibitor (barnstar)"
 /codon_start=1
 /transl_table=11
 /product="ribonuclease inhibitor"
 /protein_id="AAB80899.1"
 /db_xref="GI:1934648"
 /translation="WRKIIIDGRDFENIEVHDDLKDLKDLPPDYGRNLDALDCLTG
 WVDPLPLTKNKEFSNTFLGSYADVLEVIQEAQELKDFEKLIIIE"
 6243. .6716
 /gene="yrdG"
 6243. .6716
 /gene="yrdG"
 /note="similar to *E.coli* leucine-responsive regulatory
 protein"
 /codon_start=1
 /transl_table=11
 /product="regulatory protein YrdG"

```

/protein id="AAB80900.1"
/db_xref="GI:1934649"
/translation="MGIMLDETDKAILRLQEDASISNLKSLKUGLSPSACIARTKN
LYEAGLKKPTIVDEKGLGIWVALALINLSPNRETHSFLIEDINKPPOVECVTL
TQSHYMLKIVADMEYSYFNFIIDSLMQPALSIGVDTISVMTEKTSVPIDEM"
6729..7493
/gene="yrdH"
CDS
6729..7493
/gene="yrdH"
/notes="similar to H.influenzae hypothetical protein
H11738"
/codon_start=1
/transl_table=11
/product="hypothetical protein yrdH"
/protein id="AAB80901.1"
/db_xref="GI:1934650"
/translation="MNKNESLHSSPAINTHMNKSQIWAAPSPAPVPIPIFAGFL
FLGIAYGIEMHSGPSAIPIPIIMFSPFAGSHFVAAFLGAPNEMNALFLTLMYNA
RHLFYGISMLDKYRGTKKKLYLIFGMCDESPSINTANVPANVKGWFFVLLLNH
LYWVAGAAIGGIFSVYKFNTEGLDFVMTALFVIFIEQWMKEKHYSALTIGLSVA
SLILFGNFIPIIPLMLAILLGVLTNRKPLEKAEVSV"
7490..7822
/gene="yrdI"
CDS
7490..7822
/gene="yrdI"
/notes="similar to H.influenzae hypothetical protein
H11737"
/codon_start=1
/transl_table=11
/product="hypothetical protein yrdI"
/protein id="AAB80902.1"
/db_xref="GI:1934651"
/translation="MTMTQOMITIAMVVLGMLTRFLPFMIFPSKPTPKYVQVYIG
KVLPSAVIGLLVIYCLKDVSLLSGSHGIPELVGAADVLLHLWKNKMLLSIAGTVVY
MVLQVLF"
7833..7856
/notes="transcription terminator"
7987..9309
/gene="yrdJ"
CDS
7987..9309
/gene="yrdJ"
/notes="similar to branched chain amino acid carrier BrnQ
from Lactobacillus delbrueckii"
/codon_start=1
/transl_table=11
/product="amino acid transporter"
/protein id="AAB80903.1"
/db_xref="GI:1934652"
/translation="MSKKVSASYIIILGLMLFALFCAGNLIPLPMLQGLAKNVVA
NAGFLVTGVLPLIATVPSGKQNSLASRHPVFGIVFTIILXIALGPFAIPR
SGNVSEIGVKPLISNDASVSLIIFTILFALCLSLNPSKIIDIVGKFLPIKLT
FISLVAVALIRPIGTITQAPSKGYSQAFFKQEGYLTDLDAVAFVFGHIIIVNALKE
Query Match 13.68; Score 162.4; DB 1; Length 17828;
Best Local Similarity 51.48; Pred No. 2.7e-35;
Matches 463; Conservative 0; Mismatches 416; Indels 21; Gaps 3;
QY 227 CTATTGAAATTCATCATTAACATGGACCGCGAAGCATACAAAATCGTTGAGTCG 286
DB 4401 CTTTACAAACATATGCTTAATTCAGATCCCTGATCACACCGCTTGGAATCTTG 4460
QY 287 TGAACAAAGCGTTTACTCGCGCGTGTGAAGAGCATGGGAACCGAGAAATTCAGAAATCA 346
DB 4461 TTCAAAAAGCATTTACGCATCGGATGATTTTGCAGTTTGAAGACAAAATCCAGCATATCG 4520
QY 347 CAGATGAACTGATTCAAAATTCAGGGCGCAGTGGATTGACCTGTTCAACGATTTT 406
DB 4521 CAGATTCTTTATTGGATAAAGTGCAGCCCAACAAATTCATGAACCTAGACAGGATATG 4580
QY 407 CATACCCGCTTCCGGTATTGCTGATATCTGAGCTGCTGGGAGTGCCTTTCAAGCAGATGG 466
DB 4581 CGTTTCCATTGCCGATATCGTGATAGTGAATGCTCGGCAATTCCTCTTGAAGATGCTC 4640

```

```

QY 467 ARCAGTTTAAAGCATGCTGTCATCTTCTGTGTCAGTACACCGAAGGATAAAAGTGAAGAG 526
DB 4641 AAAAAATTCAGAGTATGGTCGCA-----ACCAATTTGATTTTTCAGATGCC 4688
QY 527 CTGAAAAAGCCCTTTTGGGAAGAACGAGATAAGTGTGAGGAAGAACTGGCCGCGTTTTT 586
DB 4689 CTGAAAGA-----CTTCAGGAATGATGCCACCTGTTAGGGAGTTTGTGTAATATCTGG 4742
QY 587 CCGGCATCATAGAAAGAAACGCGAAACAAACCGGAACAGGATATATTTCTATTATTAGTGG 646
DB 4743 ATCTCTTAGTTTCGCAAAAAAAGGAGAGAACCTCAGGAGATTTAATCAGTGCATTAATTC 4802
QY 647 AAGCGGAAGAAACAGCGGAGAGAGTGTCCGGTGAAGAGCTGATTCGGTTTTTGACGCTGC 706
DB 4803 AAGCGGAAGTGAAGGACACAAATTAAGTACAGAGAAATGTACTCGATGATATTATGCTGC 4862
QY 707 TGTGTGGTGGCGGAAATGAACCACTACAAACCTGATTTTCAAAATGCGATGTACAGCAAT 766
DB 4863 TTATCGTTGCAGTTCATGAACCAACCGTAACTTAATCAACAATATGACATATGCTCTGA 4922
QY 767 TAGAAACCGCAGCGCTTTAGCAGGAACTGGCGAGCCATCTGAACTGATGCTCCTCAGCAG 826
DB 4923 TGTGTCAACATGATCACTAATAAAGCTCGCGCAGCAGCCGAGCTTGAATGAATTCAGCCA 4982
QY 827 TGGAGGAAGCTTCGCTTTTCAGAGCGCGCGCCCGTTTTTGAGCGCATTCGCCAAGCGGG 886
DB 4983 TAGAAGGAAGCATTTGCGGTTTTCAAGCCCTGTTGAGCTTAAACGAAATTCGCTGGAGCGCTG 5042
QY 887 ATACGAGAGATCGGG---GGGCACCTGATTAAGAGGTGATATGTTTGGCGTTTGTGG 943
DB 5043 AACCGTTTATCTGATGCGGCGGAGGATTAACGGAAGATGTGATCATATTTCTTTAG 5102
QY 944 CATCGGCAAAATCGTGATGAAGCAAAAGTTTACAGACCGCAGCATGTTTGATATCGCGGCC 1003
DB 5103 CATCAGCAAAACCGTGATGAAGAAATATTTTCAAAATGCGGCATATTTGATTAAGAAAGAA 5162
QY 1004 ATCCCAATCCGCATATTGCGTTTGGCCAGCGCATCTCAATTTTGTCTTGGGCGCCGCTTG 1063
DB 5163 AGAACACCGCCATTTGCTTTTGGCCATGGCAATCATTTTGTCTTGGTGCACAGTAG 5222
QY 1064 CCCCTCTTGAAGCAAAATATCGGTTAAACGTTTGAATTTCTGCTTTTCTCTCATATGAGT 1123
DB 5223 CCCGCTTAGAGCAAAAGATGCGATTTCCACCTTATACGCGCGTGTCCAAATATACAGT 5282

RESULT 12
BSUB0014/c 199173 bp DNA linear BCT 07-JUL-2003
LOCUS Bacillus subtilis complete genome (section 14 of 21): from 2613658
to 2812830.
DEFINITION Z991117 ALJ009126
VERSION Z991117.2 GI:32468796
SOURCE
ORGANISM Bacillus subtilis subsp. subtilis str. 168
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 199173)
REFERENCE
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Etlan, K.D., Errington, J., Fabret, C., Ferrati, E.,
Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J.,
Grandi, G., Guiseppe, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R.,
Henaux, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,

```

Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetelle, D., Porwollik, S., Prescott, A.M., Prescann, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, P., Sekiguchi, J., Sekowska, A., Seror, S.J., Serrero, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipac, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.

The complete genome sequence of the gram-positive bacterium

TITLE
Bacillus subtilis
 Nature 390 (6657), 249-256 (1997)
JOURNAL
 MEDLINE
 PUBMED
 98044033
 9384377
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 199173)
 Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
 Direct Submission
 Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur,
 Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724
 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
 adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
 68 89 48

COMMENT
 On Jul 7, 2003 this sequence version replaced gi:2634966.

This entry contains data from release R16.1 of the Subtilist
 database. Further data on gene annotation and detailed information
 about changes from previous releases can be found at
<http://genolist.pasteur.fr/Subtilist/>.

FEATURES
 source
 1. 199173
 organism="Bacillus subtilis subsp. subtilis str. 168"
 /mol_type="genomic DNA"
 /strain="168"
 /db_xref="taxon:224308"
 complement(71..1030)
 /gene="phoH"
 /locus_tag="BSU25340"
 complement(71..1030)
 /gene="phoH"
 /locus_tag="BSU25340"
 /note="alternate gene name: yqfE"
 /codon_start=1
 /transl_table=11
 /product="phosphate starvation-induced protein"
 /protein_id="CAB14476.1"
 /db_xref="GI:2634980"
 /db_xref="GOA:P46343"
 /db_xref="SWISS-PROT:P46343"
 /translation="MTEHLAMQKLNPKDEALSLFGNDSFLKMKELDNLNITRG
 ETIVSGDSFQIADRLGSLALIRKGIETSERDVIYAIKWAKNLEYESMVEE
 EITENAKSKIRVTKQREYVAMKNDLVPICGACGKTYLAVKVAHLKNGHI
 KXILITPAVAGESLGPLGDKKVDPIYRPLDHLVIGADHRLMERGLIEI
 ALPMRGRTIDAYDILVDAQNTPAQMKPLTIFGSSKMIITGDSVLDLPKGVK
 SGLAVKMLKIGISMIELDQTDVVRHPLVAKIIIEAYDKN"
 complement(1027..2223)
 /gene="yqfD"
 /locus_tag="BSU25350"
 complement(1027..2223)
 /gene="yqfD"
 /locus_tag="BSU25350"
 /function="unknown"
 /note="similar to stage IV sporulation protein"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB14477.1"
 /db_xref="GI:2634981"
 /db_xref="GOA:P54469"
 /db_xref="SWISS-PROT:P54469"
 /translation="MKNKWLSFFSGKVLQELTGRGIERLLNCTKQIPVFRVKKKE

AVSLIQLQDVHAFRRVSKPKKFRINRKGFPFLLKSKLNIGFTIGFAIFFILLF
 LLSNVKIDVTGKAPDETEHQHROHLNIGVKKGRLOPLMSPKIQKSLTWGIDKNT
 WVGDLKGTTHMKVVERNEPEKIVSPRNIVAKKATITMSVQKQGFMAIHVHV
 KQGLLVGLIGSDHQDEKASKAIYGETWYRSEVTVPLETLENYVTKRRTKHLKS
 FGLAIPWGMTFKKEELKHPTEQEKHSLHFLGFKLPVSYKQEQTRSEEARLKYTK
 BEAVQGIKLGKQDVDEKIGENGEVSKVKLVHQTIVENGKVKLIILYQVIEDIVQTPEI
 VRETEE"
 complement(2242..2523)
 /gene="yqfC"
 /locus_tag="BSU25360"
 complement(2242..2523)
 /gene="yqfC"
 /locus_tag="BSU25360"
 /function="unknown"
 /note="similar to unknown proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB14478.1"
 /db_xref="GI:2634982"
 /db_xref="SWISS-PROT:P54468"
 /translation="MQQRNRMKAWLTRALEIPDPVMDLPRTVGRHLHIYENRGG
 LILFSEVRLMKQCCIISGNFVFKAILPEILLEGITDVRYVES"
 complement(2580..2999)
 /gene="yqfB"
 /locus_tag="BSU25370"
 complement(2580..2999)
 /gene="yqfB"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB14479.1"
 /db_xref="GI:2634983"
 /db_xref="GOA:P54467"
 /db_xref="SWISS-PROT:P54467"
 /translation="MEDLLTNLIILAIIGISAIIGKSKKEKONSQKRRKPKQHVQS
 ASPQKKQSKEDAPAPINPMVOAREASERRETARNLKLGLERDLAAAKQTVYTKQK
 MLQVNRKLVVQGVILGEVFGPPRAKKPHRTMPARK"
 complement(3024..4019)
 /gene="yqfA"
 /locus_tag="BSU25380"
 complement(3024..4019)
 /gene="yqfA"
 /function="unknown"
 /note="similar to unknown proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB14480.1"
 /db_xref="GI:2634984"
 /db_xref="GOA:P54466"
 /db_xref="SWISS-PROT:P54466"
 /translation="MDPSTMIILAVAVAILVAVFTFVPMVMTSALAAGVKISIF
 TLVGRRLRVIPNRVNPPLIKAKAGLVNQLSHYLAGNVDRVVALIAAQRAN
 IELTERCAAIDLAGRDVLEAVQMSVNPVETPTTAGVMDGIEVKAKRITVRANI
 ERLVGADEETIVAVUGGIVSTIGSSDNHKKVLENPDMISQTVLGKGLDSGTAFIL
 SIDIDVQKIGKIGALLQDQADKRIQAQAKERRAMAVAOQEMKRVEMRAKV
 VBAEAVPLAAEALFEGNIGVYMNIIKNIDATEMRDSFGKLTAKDPSDEORKS"
 complement(4041..5354)
 /gene="yqez"
 /locus_tag="BSU25390"
 complement(4041..5354)
 /gene="yqez"
 /locus_tag="BSU25390"
 /function="unknown"
 /note="similar to unknown proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB14481.1"
 /db_xref="GI:2634985"
 /db_xref="GOA:P54465"
 /db_xref="SWISS-PROT:P54465"
 /translation="MLQIKGFRAALLGIFLLSLGVQLNAKAEKQTVYVIPVEKNVEQ

GLASFTSRSLQDAKADHADHIIIDINTPGGLVKSADIMADLIITSEBIPVTAYVVKRAL
 SAGAYTALQADHIYAPGGKMGAAAIYDQGNADQKQAOQLWLAEMDEDAVKNRDPK
 YALAMADPDIDAKEVPGKGLDNLADKAIYEVGSEGTADNLSTLVKLIAGFEKAQIS
 YAKESFAEKARLWNTVPIVPIILITIAFLGTLVELSPGVGLPGTAGLIALILFFYGH
 LAAGLAGYETVLLPIAGVILILILEIPUPGGIIGLIGLAIIASLFLAAGSFTVMASL
 LIASVISTAFILITLTVLGRMKPEFKKILINDSTESGVVSNQTRDLMKVGILITF
 PLPSGTVIDDERLDVSRSGSFTFKKKVKKVIVKESRIVUREI"
 complement (5485..5931)
 /gene="ygey"
 /locus_tag="BSU25400"
 complement (5485..5931)
 /gene="ygey"
 /locus_tag="BSU25400"
 /functions="unknown"
 /note="similar to unknown proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB14482.1"
 /db_xref="GI:2634986"
 /db_xref="SWISS-PROT:P54464"
 /translation="MSLLRLNQMKLYMKNREKDKLTVVVMVKASLQNRKILKNDK
 LTEDLTLVLSRLKQKDSLOEFSNARLDLDVKQKLDILEVYLPEQLSREELT
 IVNETAEVGSASSKADMGKVMGAIKMPKVGKADGSLINKLVSSLS"
 complement (5946..6119)
 /gene="rpsU"
 /locus_tag="BSU25410"
 complement (5946..6119)
 /gene="rpsU"

Query Match 13.68; Score 162.4; DB 1; Length 199173;
 Best Local Similarity 51.48; Pred. No. 4.6e-35;
 Matches 463; Conservative 0; Mismatches 416; Indels 21; Gaps 3;

QY 227 CTATTGGAATTCATCATTTAAACATGACCGCGCGAAGCATACAAAATCGTTCACTG 286
 Db 117637 CTTTAAACAAACATATGCTTAAATTCAGATCCCCCTGATCAACCGCTTGCGAACTCTTG 117578

QY 287 TGAACAAAGCTTTACTCCCGCGGTGATGAGCATGCGAATGGACGAGATTCAGAAATCA 346
 Db 117577 TCCAAAAGCATTTACGCATCGGATGATTTTGCAGTTTGGAGACAAAATCAGCATATCG 117518

QY 347 CAGATGAAGCTGATTCAAAAATTTCCAGCGCGCAGTGGATTTGACCTTGTTCCACATTTT 406
 Db 117517 CAGATCTTTATTGGATTAAGTCGAGCGCCCAACAAATTCATGAACCTAGTAGACATTTAG 117458

QY 407 CATACCGCTTCGCGTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCACGCGAGATGG 466
 Db 117457 CGTTTCCATGCGCATTTATCGTATAAGTGAATGCTCGGCATTCCTCTTGAAGATCGTC 117398

QY 467 AACAGTTTAAGCATGCTCTGATCTTCGTGTCAGTACACGAGGATAAAGTGAAGAG 526
 Db 117397 AAAAATTCAGAGTATGTCGCA-----AGCAATTTATGATTTTTTCAGATGCC 117350

QY 527 CTGAAAAGCGCTTTTGGAAAGACGAGATAAGTGTGAGGAAGACTGCGCGCTTTTGG 586
 Db 117349 CTGAAAGA-----CTTCAGAGAAATGACCACTTTGTAGGAGTTTGTGAATATCTGG 117296

QY 587 CCGGCATCATAGAGAAAAGCGAAACAAACCGGAACAGGATATTATTTCTATTATTAGTG 646
 Db 117295 AATCTCTAGTTGCGAAAAGGAGAGAACCTGCGAGGAGATTTAATCAGTGCATTAATTC 117236

QY 647 AAGCGGAGAAACAGGCGAGAGTGTCCGCTGAGAGCTGATTCGTTTGGACGTGC 706
 Db 117235 AAGCGGAAAGTGAAGGACACCAATTAAGTACAGAGAAATTTACTCGATGATTAATGCTGC 117176

QY 707 TGTGTTGGCGCGGAAATGAACACCACTCAAACTGATTTCAAATGCGGATGACAGCATAT 766
 Db 117175 TTATCGTTGAGGTCATGAACACACCGTAACTTAATCAACCAATATGACATATGCTCTGA 117116

QY 767 TAGAAACGCCAGCGGTTTACGAGGAATCGCGACGCCATCTCTGAAGCTGCTCAGGACG 826
 Db 117115 TGTGTACCATGATCAACTAGAAAAGCTCCGCCAGCGCGGCTTGATGAATTCAGGCA 117056

QY 827 TGGAGGAAGCCTTTCGCTTTAGAGCGCGCGCGCTTTTGGAGCGCATTCGCCAGCGG 886
 Db 117055 TAGAAGAAGCATTCGCGTTTACAGCCCTGTTGAGCTTAACCAATTCGCTGGAGCGGTG 116996

QY 887 ATACGAGAGATCGGG--GGGCACCTGATTAAGAAGAGTGATGCTTTGGCGTTTGTGG 943
 Db 116995 AACGTTTATACTGCATGGCGGAGGATTAACCGGAAGAATGTCATCATCATTTCTTTAG 116936

QY 944 CATCGCAAAATCGTGATGAAGCAAAAGTTTGACAGACCGCATGTTTGATATCGCCGCC 1003
 Db 116935 CATCAGCAAAACCGTGATGAAAAGATATTTCCAAATGCGCATATTTGATATGAAGAA 116876

QY 1004 ATCCCAATCCGCATATTGCGTTTGGCAACGCATCCATTTTGGCTTGGGCGCGCTTG 1063
 Db 116875 AGAACCAACCGCATATTGCTTTTGGCCATGCAATCATTTTGTCTTGGTGACAGCTAG 116816

QY 1064 CCGCTCTTGAAGCAATATCGCTTAAACGCTTTTGAATTTCTGCTTTTCTCATATGGAGT 1123
 Db 116815 CCGCTTAGAAGCAAAAGATTCGGAITTCACCTTATAGCGCGCTGTCCTCATATACAGT 116756

RESULT 13
 AE017006
 LOCUS 300893 bp DNA linear BCT 16-MAY-2003
 DEFINITION Bacillus cereus ATCC 14579 section 9 of 18 of the complete genome.
 ACCESSION AE017006 AE016877
 VERSION AE017006.1 GI:29896139
 KEYWORDS
 SOURCE
 ORGANISM
 Bacillus cereus ATCC 14579
 Bacillus cereus ATCC 14579
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
 cereus group.
 REFERENCE
 1 (bases 1 to 300893)
 Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B.,
 Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N.,
 Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M.,
 Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M.,
 Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.
 Genome sequence of Bacillus cereus and comparative analysis with
 Bacillus anthracis
 Nature 423 (6935), 87-91 (2003)
 JOURNAL
 MEDLINE 22608415
 PUBMED 12721630
 REFERENCE
 2 (bases 1 to 303893)
 Candelon,B., Galloux,K., Ehrlich,D.S. and Sorokin,A.
 The number of ribosomal RNA operons in Bacillus cereus
 Unpublished
 JOURNAL
 REFERENCE
 3 (bases 1 to 300893)
 Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B.,
 Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N.,
 Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M.,
 Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M.,
 Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.
 Direct Submission
 Submitted (12-MAR-2003) INRA, Genetique Microbienne, Domaine de
 Vilvert, Jouy en Josas 78352, France
 FEATURES
 source
 location/Qualifiers
 join(1..105823,144296..300893)
 organism="Bacillus cereus ATCC 14579"
 /mol_type="genomic DNA"
 /strain="ATCC 14579"
 /db_xref="ATCC:14579"
 /db_xref="taxon:226900"
 /focus
 .05824..144295
 /organism="Bacteriophage phiBC6A52"
 /proviral
 /mol_type="genomic DNA"
 /db_xref="taxon:225726"
 /note="identified by comparison with Bacillus anthracis"
 192..896
 /locus_tag="BC2457"
 192..896
 gene
 CDS

/locus_tag="BC2457"
 /EC_number="2.7.8.-"
 /codon_start=1
 /transl_table=11
 /product="4'-phosphopantetheinyl transferase"
 /protein_id="AAP09420.1"
 /db_xref="GI:29896140"
 /translation="MKIPKNTLSHFKLSNLVSGKKRMKRLNSCDINRTLIGDL
 LIRSLICQYKINNEIRFLVNEYGKPVENFPHNLSSGEMVCTTANFNGID
 IERVSEIRAFKLAKFEFSAERFYDLNNNSDEQINYFYDLATLKEYSIKTKIGKLYTP
 LNSFSIKKESRLISYKHIPKPNFYKQYNIDPNTKLSACATRDFFPQBIINDIYTC
 QTIYKFSKEKINAED"
 883..1605
 /locus_tag="BC2458"
 883..1605
 /locus_tag="BC2458"
 /EC_number="3.1.2.-"
 /codon_start=1
 /transl_table=11
 /product="Thioesterase"
 /protein_id="AAP09421.1"
 /db_xref="GI:29896141"
 /translation="MQTKLFCFPHAGSASFYAKWKHFNPIVVPVELAGCYRI
 SENLYQNEERAVNDYKIMQIDDSYILFHSNGSLIAYEVARKIKDSKNVSEFL
 VLSGRNHNIRNIRHLSNEQFREVIANMGSTPSGVLSSEELMEI FLPIIRADPKI
 VETYIHDKNTQPCDIDFLIPNGENDEFTYDQIKWERYTSKTCFHSFEGNHFFNE
 NIEETANSIKRKLDSKRLNSF"
 1798..1868
 /locus_tag="BC2459"
 1798..1868
 /locus_tag="BC2459"
 /product="tRNA-OTHER"
 complement(3310..3900)
 /locus_tag="BC2460"
 complement(3310..3900)
 /locus_tag="BC2460"
 /codon_start=1
 /transl_table=11
 /product="Succinodiglycan biosynthesis protein"
 /protein_id="AAP09422.1"
 /db_xref="GI:29896142"
 /translation="MIATFGLWEAADVQONFDNSTATEAMQSLYSVWVTEELBQLF
 TYMKQEKGPFLTAGDMLFLYKSSFSRYAKWLQKYNPEVSEPTAVSELIELD
 RYNNKNTYPPYDRFMKIQPVINKPKVRMFIQNHKSELIQVTPHTYDVKVRGGISS
 PNLTDYSPIRQKQAQNFAMLTENQYGNKLAPFNK"
 4397..5044
 /locus_tag="BC2461"
 4397..5044
 /locus_tag="BC2461"
 /codon_start=1
 /transl_table=11
 /product="Antigen"
 /protein_id="AAP09423.1"
 /db_xref="GI:29896143"
 /translation="MKDTSKQKIIKVFLLSLGLVLIIGLWLPNFKNTIQNKAQATE
 KRLVQESTLKKELKNGEKTPIILGIMYQESRGECDPMQSSSELGLKNEIQET
 SLISIQGVKHFARKYKGTQKDVSDMTIQSYNMGSGYIDFVASQVQKQSDSAKFF
 SKMKYQDPAMYTCGGNKNFRYPYCYGDFYATKVNKTKLIEELLRNHSSSK"
 5362..5868
 /locus_tag="BC2462"
 5362..5868
 /locus_tag="BC2462"
 /EC_number="2.3.1.-"
 /codon_start=1
 /transl_table=11
 /product="Ribosomal-protein-serine acetyltransferase"
 /protein_id="AAP09424.1"
 /db_xref="GI:29896144"
 /translation="MFIKTERLIRKFEKMQAVHEYTSIDINVMKYIPGVTEBNT
 RNFVNENIGENAKNFVVLHKDILLDHIVFHKYFGEHTYELGWFPNPKYKNGYASE
 AQAATLEYGFKEKHLRIITATQCPQNTSYRVNMEKIGMRREGYFRKCI PHGNBWDEY
 YVAILEEE"
 complement(6235..6420)

/locus_tag="BC2463"
 complement(6235..6420)
 /locus_tag="BC2463"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAP09425.1"
 /db_xref="GI:29896145"
 /translation="MQYICIRKCVIIIVSAYNNKVIKIYIILRKGENNMETAVLAK
 FIGAIPIQISLLKVFEM"
 7321..8772
 /locus_tag="BC2464"
 7321..8772
 /locus_tag="BC2464"
 /EC_number="3.2.1.-"
 /codon_start=1
 /transl_table=11
 /product="S-layer protein / Peptidoglycan
 endo-beta-N-acetylglucosaminidase"
 /protein_id="AAP09426.1"
 /db_xref="GI:29896146"
 /translation="MINTWVEVFLMKHKLJATGILAGAILSYSSNIIATHKFPDVP
 TWADKSNVLDKQVNGYVDGTFGSNDLSDRASATKIMTKVLGKIDPNAKPSPTDS
 QNHATPPIAAAEKAGIVKSGNGIENPSGKVTTRAAMATMLVNAKLGSTANHEQVR
 FEDLGHGKEGHANLIDLKISNGHENGQPNRFTIRAEAGLTAKTMDLQONLDEK
 EVTATSVEDLNLTVASKLTAQEDISFAKIHSDSPFGHGQGFNAQNYGVSAHYL
 RAHALLESYGKSEIAYQKHNLFGLRAYDGPFKAKYLPSPYSDSYANNVYRERYL
 RESGMVYNGSTLTGMNVYASDKWAKIAGIMERIKPFRVEDYTYAKKLKPKNPETLD
 VDALSNEIPYKMYADGSSNNVSSATYYQVPYFNLKIKRSPDVAVEENKVGTVTPGT
 TIFIYREDPNGWVEFSFANGKRYWTLKNKLSM"
 9888..10085
 /locus_tag="BC2465"
 9888..10085
 /locus_tag="BC2465"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAP09427.1"
 /db_xref="GI:29896147"
 /translation="MISLGINILVPLSPFFIGGMATDPSGSTMDFWEVFLFIQIPFF
 PLVLSLVMLVRRKKEKVVH"
 11619..12941
 /locus_tag="BC2466"
 11619..12941
 /locus_tag="BC2466"
 /EC_number="3.4.24.3"
 /codon_start=1
 /transl_table=11
 /product="Microbial collagenase"
 /protein_id="AAP09428.1"
 /db_xref="GI:29896148"
 /translation="MYQNERLTWFQEGNAEFFAGSTRTNVVPKRKSIISLSSDPSAR
 YTAQOT-FAKYGSDPYKYPALOSYLYNHQPTFDKQLCLIRANDVXNDVARELLS
 KNTQNAEYQAYMQLTDNQKXNVHQVNDYLIQHPKPLAEKNEIYDVANI KDAK
 ITKHSQFNTSVEGTYYGCSKGESEDMKTSKQVNTLSQSKQSGCYTAY
 FVNYRVNAAFEYDIDVFGHVAETHEKKTITVNMAGPSGCVINEIIOHNSGTTSEN
 GKVIYLNWFGDGTISKTANPHTIYGEKSTVELTVKOSRSGKEIEQTKVTVAQDPQ
 TGSILEEKLIPENTLAKNLTPTDQTVTNVTPDKEVEISVNVNEQNTGMTWLYH
 ESDMNYVACGEDEGNTIKGKFAKPGKYYLVNVEKDDKNGEYSLLVK"
 13602..14639
 /locus_tag="BC2467"
 13602..14639
 /locus_tag="BC2467"
 /EC_number="3.4.16.4"
 /codon_start=1
 /transl_table=11
 /product="D-alanyl-D-alanine carboxypeptidase"
 /protein_id="AAP09429.1"
 /db_xref="GI:29896149"
 /translation="MNNTYNEKTHTLIKOLLKFSKVPKPGFAYWASFDSGVTYKGAAG
 LASILENIPITTKSVENIASVSKQFTAFSILLLEQGRSLSDDSIVKVFPSIGTYAEP
 VTLRH:LIHTTGLVDYMLAHAEINIEFTDILTVEQLKHLNGHQLAREFVGTKEFSN
 TGYFLLSLVVEKVSGLRQFAKERIFDSLNMKDITIVDCYPTTIIISARGYSKNEQA

gene		Query Match		13.4%; Score 159.6; DB 1; Length 300893;	
		Best Local Similarity		51.6%; Pred. No. 3.2e-34;	
		Matches 453; Conservative		0; Mismatches 404; Indels 21; Gaps 3;	
YMYVESPWEHTGDGAVHATVEDLVKNGENLTGTGGKELVVRMSEIGPELSPRETI IANEDYAFGLRLAEGFNCRYLEHSGWAGYRSYFMRPFREYLTIVVVLVSNYDGFUSKKY ANEIAGIILEK* 14889..16856					
QY	241	ATCATTAAATGAGCCCGCGGAGCATACAAATAATCCGTTTCAGTGTGAAACAAAGCCTTT	300		
DB	151930	ATGCTAAATTCAGATCCACCTAACCAACAGTCGTTTACGATCTTTAGTTCAAAAGCCTTT	151989		
QY	301	ACTCGCGCGTGTGAAGCAATGGGAACCGAGAATTCAGAAATCAGAGATGAATGATT	360		
DB	151990	ACACCGAAGATGATTGCACAAATTAGACGGAAGAAATTCAGAAATAGCAGATGATTGATA	152049		
QY	361	CAAAAATTCAGGGCGCAGTAGTGTGACCTTGTTCAGGATTTTTCATACCGCTTCGG	420		
DB	152050	AGTGATATAGCGCAAAAGGTACATTAATCTGTGGATGATTTATTCATTTCCATTACCA	152109		
QY	421	GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGCAACAGTTTAAAGCA	480		
DB	152110	ATTATGTAATAGCAGATGCTCGTATTCAAAGAGAGATCAAGCGAAATTTAGGATT	152169		
QY	481	TGGTCTGATCTTCTGGTCAGTACACCGAAGGATAAAGTGAAGAGCTGAAAAGCCTTT	540		
DB	152170	TGGTCTCATGCT--GTTATTGTCATCACCAGAAACACCTGAAGAA-----	152211		
QY	541	TTGGAGACGAGATAGTGTGAGGAGAGAACTGGCGCGGTTTTTTTCCGGGTCATAGAA	600		
DB	152212	ATAAAGAGAGCTGAAAAACAATCTGAAATTTATACATATCTTCAATATTAGTTGAT	152271		
QY	601	GAAAGCGAAACAAACCGGACAGGATATTTCTATTTTAGTGAAGCGAAGAAACA	660		
DB	152272	ATTAAGAAGAAAGCCAAAGAGAGACTTGGTAGTGCTTTTAATCTTCAGAGAGTGAA	152331		
QY	661	GCGGAGAAGCTGTCGCGTGAAGAGCTGATTCGCTTTTGCACGCTGCTGTGGTCGCGGA	720		
DB	152332	GGGCATAAATCTAGCGCTCGGGAATATATTCAAATGATATGCTATTATTTGTCGAGGA	152391		
QY	721	AATGAACCACTACAACCTGATTTCAATGGATGATACAGCATATTGAAGACGCCAGGC	780		
DB	152392	CATGAGACGACAGTGAATTTAATTAACAATAACGGTATTAGCACCTTCTTGAAATCCAAAT	152451		
QY	781	GTTTACAGGAACTGCGCAGCATCTCAATGATGCTCAGCGAGTGGAGAGCCTTG	840		
DB	152452	CAATTCAGTTATTAAAGATAATCCAAACTAATTTGATTCGGCTATTGAGGAGAGATTG	152511		
QY	841	CGTTTCAGAGCGCGCGCCCGTTTGAAGCGCATTTGCCAAGCGGGATAC---GGAGATC	897		
DB	152512	CGTTATTATTCTCCAGTTGAGGTTACAATGCAAGATGGCGAGCGAACCTTTTCAAATTT	152571		
QY	898	GGGGGCGACCTGATTAAGAAGGTGATATGTTTTGGCGCTTTGTCGATCGCGCAATCGT	957		
DB	152572	CACCATCAACAATACAGAAAGAGATATGTTTATTATTGCAATTGGCTTTCAGCGAACCGT	152631		
QY	958	GATGAAGCAAGTTTGAAGACCGCACATGTTTGATATCCGCGCCATCCCAATCGCAT	1017		
DB	152632	GATGAACAGTATTTTGAAATTCAGAAATATTTGATATTACAGGGAGAACACCGTCAAC	152691		
QY	1018	ATTGCGTTTGGCCACGGCATCCATTTTTCGCTTTGGGGCGCCCGCTTTCGCCGCTTTCGAAGCA	1077		
DB	152692	ATTGCGCTTGGTCATGCTAGTCAITTTCTGCTTAGGAGCTCCACTTTCGAGGTTAGAAGCA	152751		
QY	1078	AATATCGGTTTAAAGCTCTTTGATTTCTGCTTTTCTCA 1115			
DB	152752	AAGATTGCTATTACTACTTTGTTTAAATGAAATGCTTGA 152789			
RESULT 14					
SFU08223					
LOCUS					
10167 bp DNA linear BCT 04-FEB-2000					

DEFINITION		Streptomycetes fradiae cytochrome P-450, dTDP-glucose synthase, dTDP-glucose dehydratase, thioesterase, TyiCVI (tyiCVI), and TyiLR (tyiLR) genes, complete cds; and unknown gene.	
ACCESSION		U08223 AF145042	
VERSION		U08223.2 GI:6849140	
KEYWORDS		Streptomycetes fradiae	
SOURCE		Streptomycetes fradiae	
ORGANISM		Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.	
REFERENCE		1 (bases 1 to 10167)	
AUTHORS		Merson-Davies, L.A. and Cundliffe, E.	
TITLE		Analysis of five tylosin biosynthetic genes from the tyi1BA region of the Streptomycetes fradiae genome	
JOURNAL		Mol. Microbiol. 13 (2), 349-355 (1994)	
MEDLINE		95075319	
PUBMED		7984112	
REFERENCE		2 (bases 7082 to 10167)	
AUTHORS		Bate, N., Butler, A.R., Gandeche, A.R. and Cundliffe, E.	
TITLE		Multiple regulatory genes in the tylosin biosynthetic cluster of Streptomycetes fradiae	
JOURNAL		Chem. Biol. 6 (9), 617-624 (1999)	
MEDLINE		99398833	
PUBMED		10467127	
REFERENCE		3 (bases 7082 to 10167)	
AUTHORS		Bate, N., Butler, A.R., Smith, I.P. and Cundliffe, E.	
TITLE		The mycarose-biosynthetic genes of Streptomycetes fradiae, producer of tylosin	
JOURNAL		Microbiology 146 (Pt 1), 139-146 (2002)	
MEDLINE		20121747	
PUBMED		10658660	
REFERENCE		4 (bases 1 to 10167)	
AUTHORS		Merson-Davies, L.A.	
TITLE		Direct Submission	
JOURNAL		Submitted (31-MAR-1994) Louise A. Merson-Davies, Biochemistry, University of Leicester, Adrian Bldg., University Road, Leicester, LE1 7RH, UK	
REFERENCE		5 (bases 1 to 10167)	
AUTHORS		Butler, A.R.	
TITLE		Direct Submission	
JOURNAL		Submitted (23-AUG-1999) Biochemistry, University of Leicester, Adrian Bldg., University Road, Leicester, LE1 7RH, UK	
REMARK		Sequence update by submitter	
COMMENT		On Feb 1, 2000 this sequence version replaced gi:473596.	
FEATURES		Location/Qualifiers	
source	1..10167	/organism="Streptomycetes fradiae"	
CDS	/mol_type="genomic DNA"	/strain="T59235"	
	/db_xref="taxon:1906"	944..2197	
	/note="ORF1"	/codon_start=1	
	/transl_table=11	/product="cytochrome P-450"	
	/protein_id="AAA21341.1"	/db_xref="GI:473597"	
	/translation="MTQTLEAEKPEPERTANSVHPLAQAPAGAGAGLLEWFAARA EAPVWDESRQAWFRYDDYITVTNPLQFSDFSPVPFVPEELAILMGPFGGID PRHGLRLKVSQATPRRIATLEPRIABITRGLLDGLREKQIDVVSGLYLPVIV IAELLGPAEDRLREWDVILNNEGMEYVNLPDFSETMGPATKEWGDYLVRIAL KRETPDDMLMSGLIENVEGRRLTDEBEIVNIVALLTAGHSATLGLNLFVLDRH NQAELRADRLPGAEETRLRYSPFNIFRLKEDTDILGHPMKAGQMVVAVIASA EDHEVEFYBEDELTAARLPVTVTRH"	2677..3843	
	/note="ORF2, GTG start codon"	/codon_start=1	
	/transl_table=11	/product="unknown"	
	/protein_id="AAA21342.1"	/db_xref="GI:473598"	
	/translation="MTGLPRPAVRVPHDLROVHAATGVSEIGGALLRAVARGRYLL		


```
GAEALAEFEERPAEYCGNAHCVAAGSGDLDDARLALWALCVGSGDEVIIVPSHTFIASWLA
VSATGATFVPEVPGSPGCAFLDPDRLEAALTPTTRAVMPVHLVGHVDPDIPVG
AFAPHGILAVDAEAQATARYRGRRIIGSGHRTAFSFPKGNLGAIDGCGAVVTDPEL
ADRLRLNRYGAREKYRHEERGTSRLDELQAAVLVSVKLPYLDANWTRRRIEAAEYGE
ALAGLPCVTPPEGSAVPVMOYVLRSPYRDLRRRLAAGVETLVHYVPAVAVHAGAY
AGAGCPAGGLPRARLAGEVLSLPIGPHLZDEAVEVIAAVQSAALDSWEEGP"
3840..4751
/notes="orf3; GTG start codon"
/codon_start=1
/transl_table=11
/product="dUDP-glucose synthase"
/protein_id="AA021343.1"
/db_xref="GI:473599"
/transl_table="MNRDPRAMKGIILAGSGTTLRLPLTCTLSKQLLVYDKMIYY
PLSVILAGIREIGIISKDLDFLSLGGDLGLSISYAEQREPRGIAEAFILGA
RHGGDDAALITGNVDFPGFSSVLGTGTVARLDGCELFYGVKDAHRYGVGEIDSGG
RLLSLEKPRPLPGRHRLVLYTNDVVEIARTISPARGELEITDVKVYLEOGRAA
HGAGAVNWLDMGTHDSILQAGOVQLLEQFQGERIACIEBIAMRMGFIASAEQYRLG
QELRSSYSGTSLIDVAMGAADSRAQ"
4812..5813
/notes="orf4"
/codon_start=1
/transl_table=11
/product="dUDP-glucose dehydratase"
/protein_id="AA021344.1"
/db_xref="GI:473600"
/transl_table="MRLVLTGAGFISGHTFQGLTGAYPDLGATRTVLVDKLTVAGN
PANLEHVAGHPDLFVFGDIADHGWRRRLMEGVGLVHFHAAESHVDRIESEAFVR"
NVECTRVLQAAVAGVGRFVHISTDEVYGSIAEGSWPEDHPVAPNSPYAATKAASDL
LALAHRTYGLDVTRTNNYGRQYPEKAVPLFTNLDLPLVPLDGNGTREMIL
HVDHCRGVALDVAGGAPGVYINIGGTTELNTABELTRILBLCADRSALRRVADRPQ
HRRYSVDITIKIBELGYAPRTGTGLEAGTVAMVDRNARWWEPLKRSPPGRELERA"
5875..6636
/notes="orf5; GTG start codon"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="AA021345.1"
/db_xref="GI:473601"
/transl_table="MMLRCYAPVPTPAHRLVCFPHAGGSARAYRPFALAEAAAGVET
HAVYQGRDRRKEPFAATLEELAEVLPELRLLDAPDGPVVALFGHSMGAVVAYET
ARLHRSGAPAGLILSGREAPATDRTEAHLLGDRLELAEIRLQGTDPGALADEB
VLMVLPAIRDGYAAVGRYRHPVGRPCPLITVFTGADPNVITPBAEMRELITGAF
ALRVFPGGHFYLNDQRAVACVTIBETLKHGSKAH"
6763..8130
/gene="tylCVI"
/notes="required for mycarose biosynthesis"
/codon_start=1
/transl_table=11
/product="tylCVI"
/protein_id="AAF29379.1"
/db_xref="GI:684914"
/transl_table="MTLPGDIRMARRRAGRGAPVVALIPFAMRGWHEFDANGNLRHE
SGRFPSEGLRVGRDGRVWGQPIIVQPEVGLIGIVAREIHLVHLEFVQAKMEPGN
RNTLQ1SPTQATSNFTGRGRIFDLFLEPGRARVLVDVLSQAEWFLAKRN
RMVVEIYGPDELETGDFRMLTQIRALLTADNLMMDARSVLACLPTAGGQVAG
DDFAALERSPTDRSADTLCSSAELGSMFTGVRALEELVORLIPLDVAHGHWHGPD
EIAHRSHFVRLVADITASNRVARNWSQPLIQBHPSLLALIYKRVVDGVLHLYQAR
VDVGHNVNVAIVTPSDFAAGSTDEDTAGDYAVMTFPGQUTELLAHNGYNNVRLTLIAC
AHALY"
complement(8270..9562)
/gene="tylP"
complement(8270..9562)
/gene="tylR"
/function="regulation of tylosin production"
/notes="tyl orf7"
/codon_start=2
/transl_table=11
/product="TylR"
```

```
/protein_id="AAF29380.1"
/db_xref="GI:6849142"
/transl_table="MSAVTGAATPVTEPWPVLPGRASVPAVRTPDAAAGDPCAAAGDICTD
VLATLRSDQRRKERYVHGLLHTEGRITERNIAAMIHEHAGEQSLHEPFISSSTDWS
LRLARLARLEOLASVGVVPMVVPKAGTQSGVRRYVPHLQTVTSNHSWGLMY
LRESGAVPVNWLSTIGDGLMDEGLRRRAAIPRELARPSEAVAAGIYVETAGWGLPR
KPLVMDARELPVASILRALSAGOPFMLINDGTTLLAPGLSGAGRPVTATARTAEEL
ARQQRPVEMWDPAEPVPTSLALLPFCMFGLLPVFGTVARPGAGAGNTERPAPR
SLVLAEMQPDPRSVVLMNDAGRTLLRLKFLRRVETGSGAGAGHDVGLRDEF
GRSYPGWHRHVTLASLAHAMVWLAAPDIPAAAGRRATA"
Query Match 13.1%; Score 156.4; DB 1; Length 10167;
Best Local Similarity 49.3%; Pred. No. 1.4e-33;
Matches 409; Conservative 0; Mismatches 421; Indels 0; Gaps 0;
QY 250 ATGACCCCGCGAAGCATACAAAATCCGTTCCAGTCGTGGAACAAAGCCCTTACTCCGCGC 309
DB 1244 ATCGACCCCGCCCGCAGCGCCGCTGCGCAAGCTGGTCTCGCAGGCGTCCACCCCGCG 1303
QY 310 GTGATGAAGCAATGGGAACCGAGAAATTCAGAAATCACAGATGAAGTCAAAAATTT 369
DB 1304 CGGATCGGCACACTCGAAACCGCGGATCGCGAGATCACCGCGGATTCGTGGAACGGGCTG 1363
QY 370 CAGGGCGCAGTGTGACCTTGTTCACGATTTTTCATACCCGCTCCGCTTATTGTG 429
DB 1364 CGGGAAGAGGCGCAGATCGAGCTCTCCGATCTGGCCCTATCCGCTGCCGCTGATCGTC 1423
QY 430 ATATCTGAGCTGCTGGGAGTGCTTTCAGCGCAGATGGAAAGTCTTAAAGCATGGTCTGAC 489
DB 1424 ATCCGCGAGCTGCTCGGCATACCGCGGAGGACCGGTGATCTGTTCCGCGAGTGGGTGAC 1483
QY 490 CTTCTGTCAGTACCGAAGGATAAAAGTGAAGAGCTGAAAAGCCCTTTTGGAGAA 549
DB 1484 GTCATCTCAACACAGAGGCGATGATATACCCCAATCTCCCGCGAGACTTCTCCGAGCG 1543
QY 550 CGAGATAGTGTGAGAAAGACTGCGCGCTTTTTCGCGGCATCATAGAAAGAAAGCA 609
DB 1544 ATGGGCCCGCCCATCAAGAGTGGGGCGACTACCTGTACCGAGCATCGCGTCAAGCGG 1603
QY 610 AACAAACCGGAACAGSATAATTTTCTATTTTAGTGAAGCGGAAGAAACAGCGCGAGAG 669
DB 1604 GAGAGCCGAGCGAGCATCTGATGAGCGGCTCATCGAGGCGAGTTCGAGGCGCGCAGG 1663
QY 670 CTGTCGCGTGAAGAGCTAATCCGTTTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729
DB 1664 CTCACCGAGAGGAGATCGTCAACATCGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1723
QY 730 ACTACAAACCTGATTTCAAAATCGGATGTACACATATTAGAAACGCCAGCGCTTTACGAG 789
DB 1724 AGCGCCACGCTGCTCGGCAACCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1783
QY 790 GAATCTGCGCAGCATCTCTGAATGATGCTTCAGGCGAGTGGAGGAAGCCCTTGCCTTTTCA 849
DB 1784 GAGCTGCTGCGCGACCGTGAACCTGATCCCGCGCGCATCGAGGAGAGCGTTCGCTACCGC 1843
QY 850 GCGCGCGCGCGGTTTTCAGCGGCAATTCGCAAGCGGATACCGAGATCGGGGCGCACCTG 909
DB 1844 AGCCCGCTTCAACAAACATCTTCGCGCTGCTGGAAGGAGGACACGAGCATCTCCGCCCAT 1903
QY 910 ATTAAAGAGGATGATATGTTTGGCGTTTGTGGCATCGGCAATCTGTCATGAAGCAAG 969
DB 1904 ATGAAGGCGCGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1963
QY 970 TTGTGACAGACCGCATGTTTGTATATCCCGCCCATCCCAATCCCGCATATTCGCTTTTGGC 1029
DB 1964 TTCTCGATCCGACACAGTTTCAGCTGCGCGCGCGACCAACAGACACATGCTGTTGCGG 2023
QY 1030 CACGCGATCCATTTTTCGCTGCGGCGCGCGCTTTCGCGCTTTCGAGCAAA 1079
DB 2024 CACGCGATCCACCACTCGCTGCGGCTGCTTCTGCGCGCGCTGAGGCGCAA 2073
```



```
/note="Possible pseudogene of M. tuberculosis paralogs PPE"
/pseudo
/codon_start=1
/transl_table=11
/product="PPE family protein (pseudogene)"
complement(5464..5649)
/gene="ML1968"
gene

/pseudo
complement(5464..5649)
/gene="ML1968"
CDS

/pseudo
/note="Possible pseudogene of M. tuberculosis paralogs PPE"
/pseudo
/codon_start=1
/transl_table=11
/product="PPE family protein (pseudogene)"
complement(5661..8043)
repeat_region

/note="Dispersed repeat, LEPRP, copy 6"
complement(5848..5946)
/gene="ML1969"
gene

/pseudo
complement(5848..5946)
/gene="ML1969"
CDS

/note="Similar to Agrobacterium tumefaciens transposase
TR:Q44454 (EMBL:Z18270) (366 aa) fasta scores: E():
0.0002, 51.5% id in 33 aa, and to Pseudomonas putida
transposase tnpal TR:Q9R9U9 (EMBL:AJ245436) (355 aa) fasta
scores: E(): 0.00088, 48.5% id in 33 aa"
/pseudo
/codon_start=1
/transl_table=11
complement(6211..7354)
/gene="ML1970"
gene

/pseudo
complement(6211..7354)
/gene="ML1970"
CDS

/note="Similar to many e.g. Cryphonectria parasitica
(Chestnut blight fungus) putative maturase TR:AF27656
(EMBL:AF218567) (778 aa) fasta scores: E(): 7.8e-11, 29.3%
id in 216 aa"
/pseudo
/codon_start=1
/transl_table=11
/product="putative group II intron maturase-related
protein"
complement(6674..6727)
/gene="ML1970"
stem_loop

complement(7559..7698)
/gene="ML1971"
gene

/pseudo
complement(7559..7698)
CDS

/gene="ML1971"
/note="Similar to Pseudomonas putida transposase tnpal
TR:Q9R9U9 (EMBL:AJ245436) (355 aa) fasta scores: E():
3.4e-05, 47.8% id in 46 aa, and to Agrobacterium
tumefaciens transposase TR:Q44454 (EMBL:Z18270) (366 aa)
fasta scores: E(): 0.00026, 41.3% id in 46 aa"
Query Match 12.6%; Score 150; DB 1; Length 342303;
Best Local Similarity 48.8%; Pred. No. 2.1e-31;
Matches 439; Conservative 0; Mismatches 455; Indels 6; Gaps 1;

Qy 238 TCATCATTAACATGACCCCGGAGCATACAAAATCCGTTTCAGTCGTGAAACAGCC 297
Db 134656 TCGTTTATGTTCTCGATCCACCATATACCCGGTTACGCAAGCTTGTGAGCAAGCG 134597

Qy 298 TTACTCCGCGGTGATGACCAATGGACCGAGATTCAGAAATCACAGATCAACTG 357
Db 134596 TTCGCACGAAAGTCGTTACGCGCGCTGGAGGTGATATTGTCGCGTGGTGGATTCCCTG 134537

Qy 358 ATTCAAAAATTCAGGGCGCGAGTGAGTTTGACCTTGTTACGATTTTTCATACCGGCTT 417
Db 134536 CTGGATAAGGGTGGCGCGGCTGGCCAGATTTCAGTCATTTGCTAGCTAGCTTCGGCTT 134477
```

```
Qy 418 CCGGTTATTGTGATATCTGAGCTGCTCGAGTGCCCTTCAGCGCAGATGGAACAGTTTAAA 477
Db 134476 GCGGTGGCTGTGATTTGTTCGGCTGCTGGGTGTGCCGTACGAGACGCGCCGAGTTCGGT 134417

Qy 478 GCATGGTCTGATCTTCTGGTCACTACA-----CCGAAAGGATAAAAGTGAAGAAAGCTGAA 531
Db 134416 CGCGTCTCAGCACATATTGGTGCAGTCAGTCAGCCGTTTACCATTTACCGGTGAACCG 134357

Qy 532 AAAGCCTTTTGGGAAGACGAGATAAGTGTGAGAGAACTGGCGCGGTTTTTGGCCGCG 591
Db 134356 CCGGAGGCCCATCTGAAGAGCGGCTCGCGGCTGTGTGTGTGCGGATTACTTGGAGCAG 134297

Qy 592 ATCATAGAGAAAGCGAAACAAACCGGAAACAGGATATTATTTCTATTATTAGTGAAGCG 651
Db 134296 CTAGTTAAATGTCACCGGTTACACCGGTGAGGATCTGATATCAGGCTGATTGAGTTG 134237

Qy 652 GAAGAAACAGGCGAGAGCTGTCCGGTGAAGAGCTGATTCGCTTTTTCGACGCTGCTGCTG 711
Db 134236 GACGAGTCGGGTGATCAGCTGACCGGAGAGAGATCATCGCAACGTCGCGTCTGTGCTC 134177

Qy 712 GTGCCCGGAAATGAACCACTACAAACCTGATTTCAAAATGCGATGTACAGCATATAGAA 771
Db 134176 GTTCTGCTCATGAGACCAACCGTGAATCTGATTCGCAATGCGTCTCTGCGATGTACGC 134117

Qy 772 ACGCCAGCGCTTTACGAGGAACTGCGGAGCCATCCTGAACCTGATGCTCAGGCACTGAG 831
Db 134116 AACCGGTGCGAGTGGAAAGCGCTAAGTAGCAACCGCGAGCGCGCTGGTGGTTGAG 134057

Qy 832 GAAGCCTTTCGCTTTCAGAGCGCGCGCGCTTTTGGAGCGCATTTGCCAAGCGGATACG 891
Db 134056 GAGACGCTTCGTTATGATCCAGCAATCCATCTAATCGGTGCGTAGCTGCTAAGGACATG 133997

Qy 892 GAGATCGGGGGCGACCTGATTTAAAGAGGTGATATGGTTTGGCGTTTGTGGCATCGCA 951
Db 133996 ACGATCGGTCAAGCAAACTTTAAACGAGGTTGATACAATGTTTGTGCTTTCGCGCGCC 133937

Qy 952 AATCGTGAATGAAGCAAGTTTGGACAGCGGCACATGTTTGATATATCCGCGCGCATCCCAAT 1011
Db 133936 AATCGGATCCAGCGGTTTACTCAGCCCCCGATGANTTTGACCCGACCGTCCAGCTCA 133877

Qy 1012 CCGCATATTGGTTTGGCCACGCGCATCCATTTTTCCTTGGGGCGCGCGCTTGGCCGCTTT 1071
Db 133876 CGACATTTGGCATTTGCGTTCGGCTCACTTTTGGCTAGAGAGCGCATTTGCTCGGTTG 133817

Qy 1072 GAAGCAATATCGCGTTAACGCTCTTTGATTTCTGCTTTTCTCTCATATGAGAGTGGCTG 1131
Db 133816 GAAGCGACTGTCACTATTATCTCGATCATGTCGCGGTTTCCACAGGTACAGCTGCGCGCT 133757
```

Search completed: May 28, 2004, 21:28:05
Job time : 3283 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 22:08:41 ; Search time 223 Seconds
(without alignments)
501.744 Million cell updates/sec

Title: US-09-869-334B-1
Perfect score: 2052
Sequence: 1 MNVLNRQALQALLNGKK.....SITPIENSVIYGLKSRVYM 396

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqpl980s.*
2: Geneseqpl990s.*
3: Geneseqpl2000s.*
4: Geneseqpl2001s.*
5: Geneseqpl2002s.*
6: Geneseqpl2003as.*
7: Geneseqpl2003bs.*
8: Geneseqpl2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2052	100.0	396	3 AAB15501	Aab15501 Bacillus
2	2042	99.5	396	3 AAB15502	Aab15502 Bacillus
3	2018	98.3	396	3 AAB15503	Aab15503 Bacillus
4	670	32.7	410	6 AAE36997	Aae36997 Micromono
5	649.5	31.7	404	6 ABR82138	Abt82138 Bacterial
6	642	31.3	416	3 AAB18654	Aab18654 Amino aci
7	642	31.3	416	3 AAY67218	Aay67218 Macrolide
8	642	31.3	416	6 ABG71678	Abg71678 S. venezu
9	642	31.3	416	6 ADA03417	Ada03417 S. venezu
10	636	31.0	425	6 ABR82128	Abt82128 Bacterial
11	629.5	30.7	415	3 AAY77196	Aay77196 S. venezu
12	626	30.5	399	6 ABP57702	Abp57702 Saccharop
13	623.5	30.4	392	6 ABR82134	Abt82134 Bacterial
14	623	30.4	396	5 ABB81315	Abb81315 Streptomy
15	606.5	29.6	398	2 AAW11585	Aaw11585 Streptomy
16	587.5	28.6	408	6 ABR82120	Abt82120 Bacterial
17	586	28.6	404	6 ABR82143	Abt82143 Bacterial
18	581	28.3	396	2 AAR60777	Aar60777 Mycinamic
19	576.5	28.1	428	7 ADD13339	Add13339 C. glutam
20	576.5	28.1	430	4 AAG90363	Aag90363 C. glutam
21	569.5	27.8	428	4 AAB79632	Aab79632 Corynebac
22	565	27.5	399	6 ABU11376	Abu11376 Protein e
23	559	27.2	403	7 ADD15221	Add15221 EryF prot
24	550	26.8	432	6 AAE34726	Aae34726 Streptomy
25	549	26.8	429	6 AAE34728	Aae34728 Streptomy

26	549	26.8	430	6	AAE34780	Aae34780 Streptomy
27	549	26.8	430	6	AAE34736	Aae34736 Streptomy
28	547	26.7	430	6	AAE34725	Aae34725 Streptomy
29	547	26.7	430	6	AAE34731	Aae34731 Streptomy
30	544	26.5	404	2	AAR14724	Aar14724 6-hydroxy
31	544	26.5	430	6	AAE34727	Aae34727 Streptomy
32	540	26.3	430	6	AAE34734	Aae34734 Streptomy
33	539	26.3	415	6	ABR82141	Abt82141 Bacterial
34	537.5	26.2	430	6	AAE34735	Aae34735 Streptomy
35	537.5	26.2	430	6	AAE34723	Aae34723 Streptomy
36	536	26.1	430	6	AAE34730	Aae34730 Streptomy
37	532.5	26.0	411	2	AAM54389	Aam54389 Actinomad
38	531	25.9	430	6	AAE34733	Aae34733 Streptomy
39	531	25.9	430	6	AAE34737	Aae34737 Streptomy
40	530	25.8	426	6	AAE34738	Aae34738 Streptomy
41	530	25.8	430	6	AAE34722	Aae34722 Streptomy
42	530	25.8	475	6	AAE34732	Aae34732 Streptomy
43	526	25.6	482	5	ABB81316	Abb81316 Streptomy
44	523	25.5	415	6	ABR82146	Abt82146 Bacterial
45	521	25.4	475	6	AAE34729	Aae34729 Streptomy

ALIGNMENTS

RESULT 1
AAB15501
ID AAB15501 standard; protein; 396 AA.

AC AAB15501;

DT 14-FEB-2001 (first entry)

DE Bacillus subtilis hydroxylating protein.

KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.

OS Bacillus subtilis.

PN WO200044886-A1.

PD 03-AUG-2000.

PF 28-JAN-2000; 2000WO-JP000472.

PR 29-JAN-1999; 99JP-00021707.

PA (KYOW) KYOWA HAKKO KOGYO KK.

XX Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;

XX WPI: 2000-548827/50.

DR N-PSDB; AAA95666.
XX New protein derived from Bacillus genus microorganism useful for producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase inhibitors.

PS Claim 8; Page 68-71; 111pp; Japanese.

CC This sequence represents a novel Bacillus derived protein having the activity of producing a hydroxylated bicyclic compound or the corresponding lactone from a bicyclic compound or the corresponding lactone. The protein is used for preparing hydroxylated compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and hypocholesterolaemic agents

XX Sequence 396 AA;

Query Match 100.0%; Score 2052; DB 3; Length 396;

Best Local Similarity 100.0%; Pred. No. 7.2e-180; Indels 0; Gaps 0;
Matches 396; Conservative 0; Mismatches 0;

QY 1 MNVLRQALQALLNGKKNQDAYHPPFMYESMRKADAPVSFDEENQVMSVFLYDDVKVY 60
 DB 1 MNVLRQALQALLNGKKNQDAYHPPFMYESMRKADAPVSFDEENQVMSVFLYDDVKVY 60
 QY 61 GKELFSSCMPQOOTSIGNSIINMDPKHTKIBSVNKAFTPRVMKOWEPRIOEITDELI 120
 DB 61 GKELFSSCMPQOOTSIGNSIINMDPKHTKIBSVNKAFTPRVMKOWEPRIOEITDELI 120
 QY 121 QKFGSRSEDLVHDFSYPLPVIIVISELLGVPSAHMEQKAWSDLLVSTPKDKSEAEKAF 180
 DB 121 QKFGSRSEDLVHDFSYPLPVIIVISELLGVPSAHMEQKAWSDLLVSTPKDKSEAEKAF 180
 QY 181 LEERDKCEBELAAFFAGIIEEKNKPEQDIISILVEAETGEKLSGEELIPFCTLLVAG 240
 DB 181 LEERDKCEBELAAFFAGIIEEKNKPEQDIISILVEAETGEKLSGEELIPFCTLLVAG 240
 QY 241 NETTTLISNAMYSIILETPGVYEEELSHPELMPQAVEEALRFPAPVLRRIAKRDEIG 300
 DB 241 NETTTLISNAMYSIILETPGVYEEELSHPELMPQAVEEALRFPAPVLRRIAKRDEIG 300
 QY 301 GHLIKEGDMLAFVASANRDEAKFDRPHMFDIRRHNPNIHAFGHGIFHFCIGAPLARLEAN 360
 DB 301 GHLIKEGDMLAFVASANRDEAKFDRPHMFDIRRHNPNIHAFGHGIFHFCIGAPLARLEAN 360
 QY 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396
 DB 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396
 RESULT 2
 AAB15502
 ID AAB15502 standard; protein; 396 AA.
 AC AAB15502;
 DT 14-FEB-2001 (first entry)
 DE Bacillus subtilis hydroxylating protein #2.
 KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
 KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
 OS Bacillus subtilis.
 OS WO200044886-A1.
 PN 03-AUG-2000.
 PF 28-JAN-2000; 2000WO-JP000472.
 PR 29-JAN-1999; 99JP-00021707.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
 XX WPI; 2000-548827/50.
 DR N-PSDB; AAA95667.
 XX New protein derived from Bacillus genus microorganism useful for
 PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
 PT inhibitors.
 PS Claim 10; Page 94-96; 11pp; Japanese.
 CC This sequence represents a novel Bacillus derived protein having the
 CC activity of producing a hydroxylated bicyclic compound or the
 CC corresponding lactone from a bicyclic compound or the corresponding
 CC lactone. The protein is used for preparing hydroxylated compounds useful
 CC as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and
 CC hypocholesterolaemic agents

SQ Sequence 396 AA;
 Query Match 99.5%; Score 2042; DB 3; Length 396;
 Best Local Similarity 99.5%; Pred. No. 6e-179;
 Matches 394; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNVLRQALQALLNGKKNQDAYHPPFMYESMRKADAPVSFDEENQVMSVFLYDDVKVY 60
 DB 1 MNVLRQALQALLNGKKNQDAYHPPFMYESMRKADAPVSFDEENQVMSVFLYDDVKVY 60
 QY 61 GKELFSSCMPQOOTSIGNSIINMDPKHTKIBSVNKAFTPRVMKOWEPRIOEITDELI 120
 DB 61 GKELFSSCMPQOOTSIGNSIINMDPKHTKIBSVNKAFTPRVMKOWEPRIOEITDELI 120
 QY 121 QKFGSRSEDLVHDFSYPLPVIIVISELLGVPSAHMEQKAWSDLLVSTPKDKSEAEKAF 180
 DB 121 QKFGSRSEDLVHDFSYPLPVIIVISELLGVPSAHMEQKAWSDLLVSTPKDKSEAEKAF 180
 QY 181 LEERDKCEBELAAFFAGIIEEKNKPEQDIISILVEAETGEKLSGEELIPFCTLLVAG 240
 DB 181 LEERDKCEBELAAFFAGIIEEKNKPEQDIISILVEAETGEKLSGEELIPFCTLLVAG 240
 QY 241 NETTTLISNAMYSIILETPGVYEEELSHPELMPQAVEEALRFPAPVLRRIAKRDEIG 300
 DB 241 NETTTLISNAMYSIILETPGVYEEELSHPELMPQAVEEALRFPAPVLRRIAKRDEIG 300
 QY 301 GHLIKEGDMLAFVASANRDEAKFDRPHMFDIRRHNPNIHAFGHGIFHFCIGAPLARLEAN 360
 DB 301 GHLIKEGDMLAFVASANRDEAKFDRPHMFDIRRHNPNIHAFGHGIFHFCIGAPLARLEAN 360
 QY 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396
 DB 361 IALTSLISAPPHMECVSITPIENSVIYGLKSPRVKM 396
 RESULT 3
 AAB15503
 ID AAB15503 standard; protein; 396 AA.
 AC AAB15503;
 DT 14-FEB-2001 (first entry)
 DE Bacillus subtilis hydroxylating protein #3.
 DE Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
 KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
 OS Bacillus subtilis.
 OS WO200044886-A1.
 PN 03-AUG-2000.
 PF 28-JAN-2000; 2000WO-JP000472.
 PR 29-JAN-1999; 99JP-00021707.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
 XX WPI; 2000-548827/50.
 DR N-PSDB; AAA95669.
 XX New protein derived from Bacillus genus microorganism useful for
 PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
 PT inhibitors.
 XX Claim 10; Page 105-107; 11pp; Japanese.
 CC This sequence represents a novel Bacillus derived protein having the
 CC activity of producing a hydroxylated bicyclic compound or the

CC corresponding lactone from a bicyclic compound or the corresponding
CC lactone. The protein is used for preparing hydroxylated compounds useful
CC as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and
CC hypocholesterolaemic agents
XX
SQ Sequence 396 AA;
Query Match 98.3%; Score 2018; DB 3; Length 396;
Best Local Similarity 98.2%; Pred. No. 9.6e-177;
Matches 389; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 MNVLRNRQALQALLNGKQDAIYHPPFMYESMRKDAVSPDEENQVMSVELYDDVKKVV 60
Db 1 MNVLRNRQALPRALLNGKQDAIYHPPFMYESMRKDAVSPDEENQVMSVELYDDVKKVV 60
QY 61 GDEKLFSSCMPQOTSIGNSIINMPDPKTKRSVVKAFTPRVKQWEPRIQEITDELI 120
Db 61 GDEKLFSSCMPQOTSIGNSIINMPDPKTKRSVVKAFTPRVKQWEPRIQEITDELI 120
QY 121 QKFGSRSEFDLVHDFSYPLPVIIVISELLGVPSAHMEQFKAWSDLVSTPKDKSEAEKAF 180
Db 121 QKFGSRSEFDLVHDFSYPLPVIIVISELLGVPSAHMEQFKAWSDLVSTPKDKSEAEKAF 180
QY 181 LEERDKCEBELAFAFAGIIEERKNKEQDIIISILVEABETGKLSGEEIIPFCTLLVAG 240
Db 181 LEERDKCEBELAFAFAGIIEERKNKEQDIIISILVEABETGKLSGEEIIPFCTLLVAG 240
QY 241 NETTTNLSNAMYSILETGVVEELSHPELMPQAVEEALRPAPAPVLRRIAKRDEIG 300
Db 241 NETTTNLSNAMYSILETGVVEELSHPELMPQAVEEALRPAPAPVLRRIAKRDEIG 300
QY 301 GHLEKGEQVLAFAVASANRDEAKFDRPHMFDIRRHENPHIAFGHIFCLGAPLARLEAN 360
Db 301 GHLEKGEQVLAFAVASANRDEAKFDRPHMFDIRRHENPHIAFGHIFCLGAPLARLEAN 360
QY 361 IALTSLISAFPHMECVSIPIENSVIYGLKSFVRKM 396
Db 361 IALTSLISAFPHMECVSIPIENSVIYGLKSFVRKM 396
RESULT 4
AAE36997
ID AAE36997 standard; protein; 410 AA.
XX
AC AAE36997;
XX
DT 23-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
DE Micromonospora carbonacea polyketide synthase (PKS) type I #3.
XX
KW Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme.
XX
CS Micromonospora carbonacea.
XX
CA2391131-A1.
XX
PN 19-NOV-2002.
XX
PD 26-JUL-2002; 2002CA-07391131.
XX
PF 26-JUL-2001; 2001US-0307629P.
XX
PR (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PA Yang X, Staffa A, Farnet CM;
XX
PI WPI; 2003-343556/33.
XX
DR N-PSDB; AAD5813.
XX
PT Novel isolated polypeptide involved in biosynthesis of macrolides by
PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT preferably for biosynthesis of rosaramicin.

XX
PS Claim 13; Page 99-101; 206pp; English.
XX
CC The invention relates to genes and proteins involved in the biosynthesis
CC of macrolides by microorganisms. In particular it relates to the nucleic
CC acids forming the biosynthetic locus for rosaramicin (a 16-member
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is
CC useful for the biosynthesis of macrolides by microorganisms. It allows
CC direct manipulation of the proteins involved in the biosynthesis of
CC chemical engineering of the proteins involved in the biosynthesis of
CC rosaramicin. It is useful to catalyse certain biochemical reactions, in
CC vitro or in vivo, to direct or enhance the synthesis or modification of a
CC polyketide, polyketide substrate or its precursor. The present sequence
CC is M. carbonacea polyketide synthase (PKS) type I protein. (Updated on 23
CC -OCT-2003 to standardise OS field)
XX
SQ Sequence 410 AA;
Query Match 32.7%; Score 670; DB 6; Length 410;
Best Local Similarity 40.1%; Pred. No. 1e-52;
Matches 143; Conservative 65; Mismatches 141; Indels 8; Gaps 2;
QY 22 DAYHPPFPYSEMEKDAVPSPDEENQVMSVELYDDVKKVGGDKELFSS-----CXPQOTS 75
Db 19 DAQQLLDWFAYMKNPVSQWETKQAHVFSYRDYQTVTTNPLIFSSDFTSVFPVPSLA 78
QY 76 SI--GNSIINMPDPKTKRSVVKAFTPRVKQWEPRIQEITDELIQFGSRSEFDLVH 133
Db 79 LLMGPGTIGGIDEPHAPLKLVSQAFTPRRIAQMLRIGQITADVLDQVQVDDQRIIDIAS 138
QY 134 DFSYPLPVIIVISELLGVPSAHMEQFKAWSDLVSTPKDKSEAEKAFLEBRDKCEELAA 193
Db 139 DLAYPLPVTIVIELLGITPKDHEKFRFWVDIILSNEGLEYENLPDDCTETVGPALIEEMSE 198
QY 194 FFAGIIIEERKNKEQDIIISILVEABETGKLSGEEIIPFCTLLVAGNETTNLSIKMY 253
Db 199 FLYAQIAHKAEPKDDILISGLCAAEVDGRKLTDEEVVNIIVALLTAGHISATLSKLEL 258
QY 254 SIETPGVVEELSHPELMPQAVEEALRPAPAPVLRRIAKRDEIGHLEKGEQVLAFA 313
Db 259 VLEEHPQAAVRAADRSLVPGVIEETLURYSPNCIFRIINEDTDILGHMPKQKQVIAW 318
QY 314 VASANRDEAKFDRPHMFDIRRHENPHIAFGHIFCLGAPLARLEANIALTSLISAF 370
Db 319 IASANRDETEVFTDPDTFDIRRESNKLHAGHGHIFCLGAPLARLEANVFLNQTLQDF 375
RESULT 5
ABR82138
ID ABR82138 standard; protein; 404 AA.
XX
AC ABR82138;
XX
DT 30-SEP-2003 (first entry)
XX
DE Bacterial P450 enzyme SEQ ID NO:40.
XX
KW Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
KW diol; alkene; chirality; thermotolerance; thermostability.
XX
OS Eubacteria.
XX
EN WO2003052050-A2.
XX
PD 26-JUN-2003.
XX
PF 05-AUG-2002; 2002WO-US024910.
XX
PA 03-AUG-2001; 2001US-0309497P.
XX
PI (DIVE-) DIVERSA CORP.
XX
PI Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short J;

XX WPI; 2003-541641/51.
 DR N-PSDB; ACP06128.

XX Novel polypeptide having P450 enzyme activity and polynucleotides
 PT encoding the polypeptide, useful for catalyzing the hydrolysis of
 PT epoxides and arene oxides to their corresponding diols.

XX Claim 104; Page 73; 365pp; English.

XX The present sequence represents a bacterial P450 enzyme. P450 enzymes can
 CC be used to catalyze the hydrolysis of epoxides and arene oxides to their
 CC corresponding diols. P450 enzymes can also be used for hydrolyzing an
 CC alkene, for producing a compound of a desired chirality, and for
 CC increasing thermostability or thermostability of a P450 polypeptide

XX Sequence 404 AA;

Query Match 31.7%; Score 649.5; DB 6; Length 404;
 Best Local Similarity 38.3%; Pred. No. 7.7e-51;
 Matches 143; Conservative 69; Mismatches 122; Indels 39; Gaps 8;

QY 26 PFPWYEMKRDAPVSPDENQ-----VMSVFLYDDVKKVGDKEPSSC-- 69
 DB 18 FYPNEADGISLADAYEAREQPGLLRVRYMAYGEPAMLATRYADARLVGDRR-PSRAEG 76
 QY 70 ----MPQOTSIGNS-IINMDPPKHTKIRSVVNKAFTRVMKQWEPRIQETIDELIQKFQ 124
 DB 77 ARHDEPQSEGRSUGILSDPPDHTRLTLVAKFTMHQVEKLRPAVRELADLIDRWV 136
 QY 125 GR-SEFLVHDVSPVPLVIVISELLGVPSAHMBOFKAWSDLLVSTPKDKSEAEKAFLE 183
 DB 137 ATGAPDLVEEFALPVPVGVICQLLGVPEVDRPFRAMSDAALST-----SSLTAE 188
 QY 184 RDCBEEELAAFPAGIIEKRNKPDQIISILVEAETGEKLSGSELIPFCTLLVAGNET 243
 DB 189 PDANQEEELRAYMRGLIEDHRAARPREDLTIGUEARDRDRUTEQBELVCLVGLVAGHET 248
 QY 244 TTNLISNAMYSLTEPGVYELRSHPELMPQAVREALFRAPAV-----LRRIAKRDT 298
 DB 249 TATQIPNFVTLDRPEQWNLREDPELVPTAVBELMRF---VPLGSGASFPRVATEDVE 305
 QY 299 IGGHLIKEGOMVLAFVASANDEAKFDRPHMFDIRRHNPNIHAFGHGIFCLGAPLABLE 358
 DB 306 VGGTLVRAGEPVLVAVGAANDPARFAPQBLDLAREGQHILGFGHGVHHCIGAPLABLE 365
 QY 359 ANIALTSLSIAFP 371
 DB 366 LQELGALLRLRP 378

RESULT 6
 AAB18654
 ID ABE18654 standard; protein; 416 AA.
 AC AAB18654;
 XX AAB18654;

22-JAN-2001 (first entry)

Amino acid sequence of picromycin/methymycin cytochrome P450 hydroxylase.

Naibonolide synthase; polyketide synthase gene; narbonolide polyketide;
 KX antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
 KX desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
 KX picromycin biosynthesis;
 KX picromycin/methymycin cytochrome P450 hydroxylase.

Streptomyces venezuelae.

US6117659-A.

12-SEP-2000.

PF 27-MAY-1999; 99US-00320878.

XX 30-APR-1997; 97US-00846247.
 PR 06-MAY-1998; 98US-00073538.
 PR 28-MAY-1998; 98US-0087080P.
 PR 28-AUG-1998; 98US-00441908.
 PR 22-SEP-1998; 98US-0100880P.
 PR 08-FEB-1999; 99US-0119139P.
 PR 20-MAY-1999; 99US-0134990P.

XX (KOSA-) KOSAN BIOSCIENCES INC.

XX Ashley G, Betlach MC, Betlach M, Tang L, Medaniel R;

DR WPI; 2000-610844/58.

PT New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
 PT for converting ketolides to antibiotics and as antibiotics and
 PT intermediates in the synthesis of compounds with pharmaceutical value.

XX Disclosure; Col 47-48; 117pp; English.

XX The present sequence represents a picromycin/methymycin cytochrome P450
 CC hydroxylase (pick). The nucleotide sequence encoding it is used in the
 CC course of the invention. The specification describes a recombinant DNA
 CC compound expressing recombinant polyketide synthase genes in host cells
 CC for the production of narbonolide, narbonolide derivatives and
 CC polyketides that are useful as antibiotics and as intermediates in the
 CC synthesis of compounds with pharmaceutical value. The DNA compounds may
 CC also encode a C12-hydroxylase (pick), desosamine biosynthesis and
 CC desosaminyl transferase enzymes (useful for conversion of ketolides to
 CC antibiotics), and the beta-glucosidase enzyme (involved in picromycin
 CC biosynthesis). These compounds are also useful for increasing the
 CC antibiotic activity of a compound relative to the unhydroxylated
 CC compound. The recombinant host cells are useful as genetic systems that
 CC allow rapid engineering of the narbonolide polyketide synthase. These
 CC would be valuable for creating novel ketolide analogs for pharmaceutical
 CC applications

XX Sequence 416 AA;

Query Match 31.3%; Score 642; DB 3; Length 416;

Best Local Similarity 37.4%; Pred. No. 3.9e-50;

Matches 148; Conservative 77; Mismatches 133; Indels 34; Gaps 12;

QY 21 QD-AVHPFPWYEMKRDAP---VSPDENQVMSVFLYDDVKKVGD---KELPSSCMP- 71

DB 22 QDFAADPYTYARLRAEGFAHRVTPGSDVWLVGYDSARAVLADPRFXDWRNSTTPL 81

QY 72 -QOTSIGNSIINMDPPKHTKIRSVVNKAFTRVMKQWEPRIQETIDELIQKF-----QGR 126

DB 82 TEAEALNHNMLSDPPRHTRLKLVAREFTMRRLPRVQEIIVGVLVADMLAAPGR 141

QY 127 SEFDLVHDVSPVPLVIVISELLGVPSAHMBOFKAWSDLLVSTPKDKSEAEKAFLEERDK 186

DB 142 A--DLMESLAWPLPTIVISELLGVPEPDRAPRVRTDAFVP--PDPAQAQTAM----- 151

QY 187 CEEELAAFPAGIIEKRNKPDQIISILVE-AETGEKLSGSELIPFCTLLVAGNETTT 245

DB 192 --AEMSGYLSRLSDSKRGQDGEOLLALVRTSDEDSRLTSEELGMAHILLVAGHETTV 249

QY 246 NLISNAMYSLTEPGVYELRSHPELMPQAVREALFRAPA-PVLRRTAKRDTGIGHLI 304

DB 250 NLLANGYALLSHPDQALALRADMTLLDGAVEMLRYEGPVESATYRFPVEVDJGTVI 309

QY 305 KEGDMVLAFVASANDEAKFDRPHMFDIRRHNPNIHAFGHGIFCLGAPLARLEANIALT 364

DB 310 PAGDTVLVLADAHHTPERFPDPHFRDIRRDTAGHLAFGHGIFCIGAPLARLEANIAVR 369

QY 365 SLISAFPMCECVSTPIE-----NSVIYGLKSFVK 395

DB 370 ALLERCPDL-ALDVSFSGELVWVFPNMRGLKALPIR 404

QY	72	-QQTSSIGNSLINMDPPKHNKIRSVNKAFTPRVMKQWEPRIQEITDELQKF----	QGR 121
DB	82	TEAEALNHNKJSSDPPHRIHLRLKLVAREFTMRVVELLRPVQELVGLVJAMLAAPDGR	141
QY	127	SEFDLVHDSYPLFVIVISELLGVPSAPMEQFKAWSDLLYSTPKDKSEEAKEAFLEERDK	186
DB	142	A--DLMESLAWPLPITVISELLGVPEPDRAAFRVMTDAFV--PDDPAQAQTAV----	191
QY	187	CEBELAAPPAGIITEKRNKPEODIISILVE-ABETGEKLSCEELIPFCTILLVAGNETTT	245
DB	192	--AEMSGYLSRLIDSKRGQGEDLLSALVRTSOEGSKRLTSEELMGMAHILLVAGHETIV	249
QY	246	NLSINSMYSILETPGVVTEELRSHPELMPQAVEALRFAPA-PVLRRAKRDTEIGGHLI	304
DB	250	NLLANGMIALSHSDQLAALRADMTLLDGAVEMLRYEGPVESATYRFPVEPVOLDGTVI	309
QY	305	KEGDWLVAFVASANRDEAKFDRPHMFOIRRHNPPIAHFGHGHFCGAPLARLENIALT	364
DB	310	PAGDTVLVLADAHRTPEFPDPHRFDIIRDTAGHLAFGHGTHFCIGAPLARLEARI	369
QY	365	SLISAFPMCEVCSITPIE-----NSVIYGLKSFVRVK	395
DB	370	ALLERCPCDL-ALFVSPGELVWYPNPMINGLKALPIR	404
RESULT 8			
ABG71678	ID ABG71678 standard; protein; 416 AA.		
XX	ABG71678;		
XX	21-JAN-2003 (first entry)		
DE	S. venezuelae picromycin/methymycin cytochrome P450 hydroxylase, Pick.		
KW	Narbonolide polyketide synthase; PKS; desosamine biosynthetic gene;		
KW	desosaminyl transferase gene; beta-glucosidase gene; antibiotic;		
KW	pick hydroxylase gene; C12 hydroxylase gene; narbonolide;		
KW	desosaminylated polyketide; narbomycin biosynthesis; Pick;		
KW	picromycin/methymycin cytochrome P450 hydroxylase;		
XX	picromycin biosynthesis; enzyme.		
OS	Streptomyces venezuelae.		
PN	W0200297062-A2.		
XX	05-DEC-2002.		
XX	22-FEB-2002; 2002WO-US305642.		
PF	22-FEB-2001; 2001US-00793708.		
ER	(KOSA-) KOSAN BIOSCIENCES INC.		
PA	Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;		
PI	WPI; 2003-041412/03.		
DR	Preparation of polyketides by recombinant DNA technology, useful as		
PT	antibiotics and as intermediates in the synthesis of pharmaceutical		
PT	compounds.		
XX	Disclosure; Page 45; 127pp; English.		
ES	The present invention relates to recombinant DNA sequences encoding for a		
CC	narbonolide polyketide synthase (PKS) domain, and methods of producing		
CC	polyketides by recombinant DNA technology. The recombinant DNA sequences		
CC	are derived from Streptomyces venezuelae desosamine biosynthetic,		
CC	desosaminyl transferase, beta-glucosidase, or pick (C12) hydroxylase		
CC	genes. The method is useful for transforming a cell with a recombinant		
CC	expression vector that encodes a functional beta-glucosidase gene, and		
CC	therefore for increasing the yield of a desosaminylated polyketide in a		

DE Bacterial P450 enzyme SEQ ID NO:20.
XX Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
KW diol; alkene; chirality; thermotolerance; thermostability.
XX Eubacteria.
OS
XX WO2003052050-A2.
XX
XX 26-JUN-2003.
XX
XX 05-AUG-2002; 2002WO-US024910.
XX
XX 03-AUG-2001; 2001US-0309497P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Weiner D, Burke M, Hitchman T, Fajol C, Richardson T, Short J;
PI
XX WPI; 2003-541641/51.
XX N-PSDB; ACF06118.
XX
XX Novel polypeptide having P450 enzyme activity and polynucleotides
PT encoding the polypeptide, useful for catalyzing the hydrolysis of
PT epoxides and arene oxides to their corresponding diols.
XX
XX Claim 104; Page 60-61; 365pp; English.
XX
XX The present sequence represents a bacterial P450 enzyme. P450 enzymes can
CC be used to catalyze the hydrolysis of epoxides and arene oxides to their
CC corresponding diols. P450 enzymes can also be used for hydrolysing an
CC alkene, for producing a compound of a desired chirality, and for
CC increasing thermotolerance or thermostability of a P450 polypeptide
XX
XX Sequence 425 AA;
SQ
Query Match 31.0%; Score 636; DB 6; Length 425;
Best Local Similarity 33.8%; Pred. No. 1.4e-49;
Matches 137; Conservative 79; Mismatches 145; Indels 44; Gaps 8;
QY 23 AYHPPWYEMRKDAV---SPDEENQVMSVFLYDDVKKVVGDKELF----- 66
DB 25 ASDPYPAYAWLRHAPVHRTLSGVEAMLVTRYGDARQALDQRLSKNFAHDESPHAK 84
QY 67 --SSCMFQQTSSIGNSIINDDPKHTKIRSVNKAFTPRVMKQWEPRIQETIDELIQK 124
DB 85 GKTGIFGERKAEMLTHLLNIDPDHTLRRLVSKAFTPRVABTPRVQBELTDLDAFV 144
QY 125 GRSEFLVHDFSYPLPVIVISELLGVPSAHMEQPKAWSIDLIV---STPKKSEAEKAF 181
DB 145 TKGSAQLIHDFAPFLPIYALICLLGVPEEDQDFRDWAGMMIRHGGPGRGVARSVK--- 201
QY 182 EERDKCEELAAFFAGFIEBKRNKPEODIISILVEAETGKLSGELIPFCTLLLVAGN 241
DB 202 -----KRGVLAELIHRKRAPGDDLSGLKASDGEHLTENAAMAFILLFAGF 253
QY 242 ETTNLISNAMYILETPGVYEELRS-----HPELMPQAVEALFRAPAPVLR-RIAKR 295
DB 254 ETTVNLGNGVYQLLRHPGQRRLQTSAAAGETGLTETGIELRLYDGPVEMATWRYATE 313
QY 296 DTBIGHLKEGDMVLAFAVANSANDEKAFDRPHEMFDTRRHPNPHIAGCHGHFCLGAPLA 355
DB 314 PLATIGGQDIPAGDPVLVLAADRDPERFDPLDRLARKNQHLGVLGHGHYCLGAPLA 373
QY 356 RLEANTALTSLSAFPMWECVSTPIE-----NSVIVGLKSFVK 395
DB 374 RLSEQTALTLTLRLPDR-LADPAELRWGRGLMRGLTLPVE 417

AC AAY77196;
XX
DT 12-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae macrolide biosynthetic enzyme PikC, SEQ ID NO:39.
XX
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae; ATCC15439.
XX
FH Key Location/Qualifiers
FH Misc-difference 103 /note= "Encoded by CCG"
PT
XX WO200000620-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-US014398.
XX
XX 26-JUN-1998; 98US-00105537.
XX
XX (MINU) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX WPI; 2000-160679/14.
XX N-PSDB; AAZB7301.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX synthesis of methymycin and pikromycin.
XX
XX Claim 19; Page 429-430; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesize methymycin,
XX pikromycin, neomethymycin, narbomycin or a combination of these
XX compounds. Recombinant or augmented cells comprising the desosamine
XX and/or macrolide biosynthetic gene clusters are useful for the production
XX of biologically active macrolides. The macrolide biosynthetic proteins
XX are useful for synthesis of methymycin, pikromycin, neomethymycin and
XX narbomycin. The alternative termination of polyketide synthesis may be
XX useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
XX monomers. The compounds produced by the recombinant host cells are useful
XX as biopolymers, e.g., in packaging or biomedical applications, to
XX engineer PHA monomer synthases or to prepare biologically active agents,
XX such as chemotherapeutics, immunosuppressants, agents to treat asthma,
XX chronic obstructive pulmonary disease as well as other diseases involving
XX respiratory inflammation, cholesterol-lowering agents or macrolide-based
XX antibiotics which are active against a variety of organisms, e.g.,
XX bacteria, including multi-drug resistant pneumococci and other
XX respiratory pathogens, as well as viral parasitic pathogens, or as crop
XX protection agents (e.g., fungicides or insecticides) via expression of
XX polyketides in plants. Sequences AAY7190-77197 represent macrolide
XX biosynthetic enzymes from Streptomyces venezuelae ATCC 15439, which are
XX encoded by sequences AMZ87295-287302. (Updated on 12-SEP-2003 to
XX standardise OS field)
XX
SQ Sequence 415 AA;
Query Match 30.7%; Score 629.5; DB 3; Length 415;
Best Local Similarity 37.4%; Pred. No. 5.5e-49;
Matches 148; Conservative 76; Mismatches 137; Indels 35; Gaps 13;

QY 21 QD-AVHPFPWYEMRKDAP---VSDEENQVNSVFLYDDVKKVGVG---KELFSSCMP- 71
DB 22 QDFAADPYTYARLARAECPAHRVTPGEGVNLVGVYDARAVLADPRFSKOWRNSTTFL 81
QY 72 -QOTSSIGNSIINMPDPKHTKIRSVNKAFTPRVMKQWEPRIQETDELQKF-----QOR 126
DB 82 TEAEALAHNMLESPPPHRTR-KLVAREFTWRRVVELLRPRVQEIVDGLVDAMLAAPDGR 140
QY 127 SEFDLVHDPSPYPLPVITVISELGVPSAHMEQKAWSDLLVSTPKDKSEAEKAFLEERDK 186
DB 141 A--DMESLAWELPITVISELGVPEPDAARVWIDAFV--PDPAQAQATM----- 190
QY 187 CSEELAAFFAGIIEKRNKPEQDIISILVE-AEETGEKLSGSELIPFCFLLVAGNETT 245
DB 191 --AENSGVLSRLDSKRGQDGLLSALVRTSDEDSRLTSEELGMAHILLVAGHETTV 248
QY 246 NLISAMYSILETPGVYELRSHPELMPQAVEALPRAPA-PVLRRIAKRTEIGCHLI 304
DB 249 NLIANGMYALLSHPDQALALRADMTLLDGAVEMLRYEGFVSATYRFPVDFVLDGTVI 308
QY 305 KEGDMVLAFVASANDEAKFDRPHMPDIRRHHPNPHIAFGHGHFCLGAPLARLEANIALT 364
DB 309 PAGDVLVVLADAHRTPEPFP2DPHFPDIRDPTAGHLAFGHGHFCLGAPLARLEARIYR 368
QY 365 SLISAFPHMECVSITPIE-----NSVIYGLKSFVX 395
DB 369 ALLERCPDL-ALDVSFGLVVMYPNPMIRGLKALPIR 403

RESULT 12
ID ABP57702 standard; protein; 399 AA.
AC ABP57702;
XX
XX
DT 22-JAN-2003 (first entry)
DE Saccharopolyspora ORF LVII butenyl-spinosyn biosynthetic gene product.
KW Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
KW metabolite; spinosyn.
OS Saccharopolyspora sp.
FN WO200279477-A2.
XX
XX
PD 10-OCT-2002.
XX
XX
PF 28-MAR-2002; 2002WO-US009968.
XX
XX
PR 30-MAR-2001; 2001US-0280175P.
XX
XX
PA (DOWC) DOW AGROSCIENCES LLC.
XX
XX
PI Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
PI Mitchell JC;
XX
XX
DR WPI: 2003-058434/05.
DR N-PSDB; ABV75558.
XX
XX
PT New butenyl-spinosyn biosynthetic genes, useful for increasing the
PT production of butenyl-spinosyn insecticidal macrolides, or for changing
PT the metabolites or products produced by spinosyn-producing
PT microorganisms.
XX
XX
PS Claim 1; Page 209-210; 218pp; English.

CC The invention relates to a novel DNA molecule comprising a DNA sequence
CC that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl
CC -spinosyn biosynthetic genes are useful for increasing the production of
CC butenyl-spinosyn insecticidal macrolides. The genes are also useful for

CC changing the metabolites or products produced by spinosyn-producing
CC microorganisms. The present sequence represents one of the butenyl-
CC spinosyn biosynthetic polypeptides of the invention
XX
SQ Sequence 399 AA;

Query Match 30.5%; Score 626; DB 6; Length 399;
Best Local Similarity 36.0%; Pred. No. 1.le-48;
Matches 140; Conservative 75; Mismatches 140; Indels 34; Gaps 8;
QY 10 LQALLGNKQDAYHPPFPWYEMRKDAPVS---FDEENQVNSVFLYDDVKKVGVGKELF 66
DB 15 IDRLLLDDR-----YAVLRGEFVKIRLPYGGGLVTRYAIKIVLIGDPRS 63
QY 67 SSCMPQQTSSIG-----NSIINMPDPKHTKIRSVNKAFTPRVMKQWEPRIQETDEL 119
DB 64 AAAILNRDVRGFPPLILREHSLGTMDPPEHTRLRKLVGKAFTARRVEQLRPRTQQLVDHL 123
QY 120 IQKFQGRS-EFDLVHDPSPYPLPVITVISELGVPSAHMEQKAWSDLLVSTPKDKSEAEK 178
DB 124 LDRMAADGPGGLVLSALALPLPKVICLLGIPVADRFRFRVMSDIALAITSNPEIR- 182
QY 179 AFLERDKCEEELAAFFAGIIEKRNKPEQDIISILVEAEETGEKLSGSELIPFCFLLV 238
DB 183 ---ESRD---QRAYIGELVQQRKMPTEDELSVLVQARAEGAQLSEEEIVVTGAGLLI 235
QY 239 AGNETTTLNISAMYSILETPGVYELRSHPELMPQAVEALPR--APAPVLRRIAKRD 296
DB 236 AGFETTANHIANFTFNLLTHPDQDLKIADPELVPAVEELLRYTTLGATPGFPRIATED 295
QY 297 TELGHLINKEGDMVLAFVASANDEAKFDRPHMPDIRRHHPNPHIAFGHGHFCLGAPLAR 356
DB 296 TELGGVSTRGDAVFEEIASANRDSAVFDGPDDELDAREHNSHMAUGHGHGPHYCIGAPLAR 355
QY 357 LEANIALTSLISAFPHMECVSITPIENS 385
DB 356 MELQVAIGTLIKRFPQLSPA--VPVDEVV 382

RESULT 13

ABR82134
ID ABR82134 standard; protein; 392 AA.
XX
XX
AC ABR82134;
XX
DT 30-SEP-2003 (first entry)
XX
DE Bacterial P450 enzyme SEQ ID NO:32.
XX
KW Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
KW diol; alkene; chirality; thermotolerance; thermostability.
OS Eubacteria.
XX
XX
PN WO2003052050-A2.
XX
XX
PD 26-JUN-2003.
XX
XX
PF 05-AUG-2002; 2002WO-US024910.
XX
XX
PR 03-AUG-2001; 2001US-0309497P.
XX
XX
PA (DIVE-) DIVERSA CORP.
XX
XX
PI Weiner D, Burke M, Hitchman T, Pajo C, Richardson T, Short J;
XX
XX
DR WPI: 2003-541641/51.
DR N-PSDB; ACF06124.
XX
XX
PT Novel polypeptide having P450 enzyme activity and polynucleotides
PT encoding the polypeptide, useful for catalyzing the hydrolysis of
PT epoxides and arene oxides to their corresponding diols.

DR WPI: 1996-097631/10.
DR N-PSDB: AAT58555.
XX
PT New streptogramin B derivs. useful as antibiotics - produced by new
PT mutants of Streptomyces having altered genes for streptogramin B
PT biosynthesis.
XX
PS Example 1; Page 113-114; 146pp; French.
XX
CC The papA gene of S.pristinaespiralis is involved in the biosynthesis of 4
CC -dimethylamino-L-phenylalanine (DMPAPA), a precursor for pristinamycin
CC IA. Upstream of the papA gene, on the complementary strand, is the snbA
CC gene coding for 3-hydroxypicolinic acid-AMP ligase. The region between
CC these two genes was sequenced and two open reading frames were
CC identified. The first (pipA) decodes to an amino acid sequence with
CC homology to ornithine cyclodeaminase from Agrobacterium tumefaciens. The
CC pipA gene product is likely to catalyze the cyclo- deamination of lysine,
CC leading to production of pipecolic acid. Mutations in the pipA gene were
CC shown to affect pipecolic acid synthesis but not the synthesis of 3-
CC hydroxypicolinic acid. The second open reading frame (snbP) could be
CC decoded to give a product with homology to hydroxylases of the cytochrome
CC P450 type. Disruption of the pipA and snbP genes can be used to produce
CC strains of S.pristinaespiralis which are unable to produce the antibiotic
CC pristinamycin I but which may be able to produce new, modified forms of
CC it
XX
SQ Sequence 398 AA;
Query Match 29.6%; Score 606.5; DB 2; Length 398;
Best Local Similarity 38.0%; Pred. No. 6, 8e-47;
Matches 136; Conservative 66; Mismatches 133; Indels 23; Gaps 7;
Qy 29 WYEMRKDAPVSFDEENQVWVVELYDDVKVVGDKELFSSCMPQ---QTSSIGNSIIN-M 84
Db 19 WLRMRHHHPVHEDEYG-AFHVFRHADVLTVASDPGVYSSQLSLRLPGSQALSEQILSVI 77
Qy 85 DPKXHTKIRSVYVVKAFTPRVKMKQWEPRIQETDELIQKFGHSEFDLVHDSYPLPVI 144
Db 78 DPFNHTLRLVVSQAFPTPTVADLEPRVTELAGQLDADVG-DTDLVADPAYPLPVI 136
Qy 145 SELGVPSSAHMEQKAWSDLLVSTPKDSEAEKAFLEERDKCEE-----LAA 193
Db 137 AELIGVPPADRTLFSSMDRL-----QMVDADPADMQFGDDADEYQRLVKEPMRAMHA 191
Qy 194 FFAGIIIEKRNKPODIIISILVEAETGKLSGELIPFCTLLLVAGNETTNNLISNAMY 253
Db 192 YLHDHVTDRRARPANDLISALVAARVEGERLTDEQIVEFGALLMAGHVSMTLIGNTVL 251
Qy 254 SILETPGVYBELSRHPELMPQAVEALFRFAPAPVIREIAKRDTEIGGHLIKEGDMVLAP 313
Db 252 CLKDHPRAAARAADRSIPALIEVLRLRPITVMAVTTKDTVLACTTIPAGRMVVP 311
Qy 314 VASANRDEAKFDRPHMFDIRRHNPPIAFGHGIFLCGLAPLARLEANIALATSLISAFP 371
Db 312 LLSANHDEQVFTDPDHLDLARE-GQIAFGHGHVCIAGPLARLEGRALAEALPDRFP 368

Search completed: May 28, 2004, 23:17:04
Job time : 225 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 18:09:49 ; Search time 400 Seconds
(without alignments)
12649.010 Million cell updates/sec

Title: US-09-869-334B-2

Perfect score: 1191

Sequence: 1 atgaatgtgttaaacgcgcg.....gttcogtgtgaaatgtaa 1191

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002s:*

7: geneseqn2003as:*

8: geneseqn2003bs:*

9: geneseqn2003cs:*

10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1189.4	99.9	1191	3	AAA95666 Bacillus
2	1186.2	99.6	1204	3	AAA95667 Bacillus
3	1181.4	99.2	1221	3	AAA95668 Bacillus
4	1168.6	98.1	1221	3	AAA95669 Bacillus
5	425	35.7	966	6	ABK74899 Bacillus
6	242	20.3	1213	6	ABK74891 Bacillus
7	149	12.5	1233	7	AAJ55813 Micromono
8	149	12.5	60196	7	AAJ55810 Micromono
9	139.2	11.7	8478	2	AAQ81792 B. subtil
10	120.4	10.1	1215	8	ACF06128 Bacterial
11	118.2	9.9	110000	4	AAI99682 39
12	116.4	9.8	110000	4	AAI99683 39
13	116.4	9.8	110000	4	AAI99682 39
14	116.4	9.8	110000	4	AAI99683 39
15	114.2	9.6	1248	8	AAH65582 Bacterial
16	112.8	9.5	1290	5	AAH65582 C glutami
17	112.8	9.5	1407	4	AAAF1749 Corynebac
18	112.8	9.5	1414	9	ADD13338 C. glutam
19	112.8	9.5	34980	5	AAH68525 C. glutam
20	111.4	9.4	1194	2	AAAF8555 Streptomy
21	111.4	9.4	1257	4	AAAF81357 Quorum se
22	111.4	9.4	4496	2	AAAF8553 Streptomy
23	110.2	9.3	1155	6	ABK74894 Bacillus

24	106	8.9	1215	8	ACF06133 Bacterial
25	105.4	8.8	8293	6	ABX09140 Mycobacte
26	105.4	8.8	110000	4	AAI99682 21
27	105.4	8.8	110000	4	AAI99683 21
28	103.6	8.7	6085	2	AAAT70153 S. longisp
29	103	8.6	215	6	ABK74951 Bacillus
30	102.8	8.6	36538	7	ABV75558 Saccharop
31	98.2	8.2	5880	4	AAI68687 Pseudomon
32	98.2	8.2	11355	4	AAI68692 Pseudomon
33	97.4	8.2	2168	2	AAQ73674 Mycinamic
34	97.2	8.2	110000	4	AAI99682 08
35	97.2	8.2	110000	4	AAI99683 08
36	96.4	8.1	1188	9	ADC36141 Weed cont
37	96.4	8.1	1418	9	ADC36144 Weed cont
38	96.4	8.1	1418	9	ADC36202 Weed cont
39	96.2	8.1	8169	2	AAV26609 Actinomad
40	96	8.1	1191	6	ABN88910 Streptomy
41	95.4	8.0	1278	8	ACF06118 Bacterial
42	93.8	7.9	1236	7	ACF69188 Phototrab
43	93.8	7.9	110000	7	ACF67367 20
44	93.8	7.9	110000	7	ACF67367 21
45	93.8	7.9	243072	7	ACF65382 Phototrab

ALIGNMENTS

RESULT 1

AAA95666

ID AAA95666 standard; DNA; 1191 BP.

AC AAA95666;

XX AAA95666;

DT 14-FEB-2001 (first entry)

DE Bacillus subtilis hydroxylating enzyme gene.

XX Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;

KW hydroxymethylglutaryl co-enzyme A reductase; hypochlolesterolaemia.

XX Bacillus subtilis.

OS WO200044886-A1.

PN 03-AUG-2000.

PD 28-JAN-2000; 2000WO-JP0000472.

PF 29-JAN-1999; 99JP-C0021707.

XX (KYOWA) KYOWA HAKKO KOGYO KK.

XX Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;

XX WPI; 2000-548827/50.

DR P-PSDB; AAB15501.

XX New protein derived from Bacillus genus microorganism useful for

PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase

PT inhibitors.

XX Claim 14; Page 64-67; 111pp; Japanese.

CC This sequence represents the coding region for a novel Bacillus derived protein having the activity of producing a hydroxylated bicyclic compound or the corresponding lactone from a bicyclic compound or the corresponding lactone. The protein is used for preparing hydroxylated compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and hypocholesterolaemic agents

XX Sequence 1191 BP; 330 A; 248 C; 311 G; 302 T; 0 U; 0 Other;

SQ Query Match 99.9%; Score 1189.4; DB 3; Length 1191;

Best Local Similarity 99.9%; Pred. No. 0; Matches 1190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	1	ATGAATGTGTTAAACCGCGGCAAGCTTTGACGAGCGCTGCTCAATGGGAAACAA 60
Db	1	ATGAATGTGTTAAACCGCGGCAAGCTTTGACGAGCGCTGCTCAATGGGAAACAA 60
QY	61	CAGGATGCGTATCATCCGTTTCATCGTATGAATCGATGAGAAAGATGCGCTGTTCC 120
Db	61	CAGGATGCGTATCATCCGTTTCATCGTATGAATCGATGAGAAAGATGCGCTGTTCC 120
QY	121	TTTCATGAAGAAACCAAGTGTGAGCGTTTCTTTATGATGATCTCAAAAAGTTGT 180
Db	121	TTTCATGAAGAAACCAAGTGTGAGCGTTTCTTTATGATGATCTCAAAAAGTTGT 180
QY	181	GGGATAAAGAGTTGTTTCCAGTTGCATCCGACGACAGCAAGCTCTATTGGAATTC 240
Db	181	GGGATAAAGAGTTGTTTCCAGTTGCATCCGACGACAGCAAGCTCTATTGGAATTC 240
QY	241	ATCATTAACNTGACCGCGGAGCAATACAAAATCGTTTCACTGCTGAACAAAGCTTT 300
Db	241	ATCATTAACNTGACCGCGGAGCAATACAAAATCGTTTCACTGCTGAACAAAGCTTT 300
QY	301	ACTCGCGCTGATGACGAATGGAACCGGAATCAAGAAATCAAGATGAAGTGAAT 360
Db	301	ACTCGCGCTGATGACGAATGGAACCGGAATCAAGAAATCAAGATGAAGTGAAT 360
QY	361	CAAAATTTACGGGCGAGTGAAGTTGACCTTTGACGATTTTTCATACCGCTTCG 420
Db	361	CAAAATTTACGGGCGAGTGAAGTTGACCTTTGACGATTTTTCATACCGCTTCG 420
QY	421	GTTATGTGATATCTGAGCTGCTGGAGTGCCTCAGCGACATGGAACAGTTAAAGCA 480
Db	421	GTTATGTGATATCTGAGCTGCTGGAGTGCCTCAGCGACATGGAACAGTTAAAGCA 480
QY	481	TGGTCTGATCTTCTGTCAGTACACGGAAGTAAAGTGAAGAACTGAAAAGCTTT 540
Db	481	TGGTCTGATCTTCTGTCAGTACACGGAAGTAAAGTGAAGAACTGAAAAGCTTT 540
QY	541	TTGGAAGAACGAGTAAAGTGTGAGAAAGTAAAGTGAAGAACTGAAAAGCTTT 600
Db	541	TTGGAAGAACGAGTAAAGTGTGAGAAAGTAAAGTGAAGAACTGAAAAGCTTT 600
QY	601	GAAAAGCAGAACACACGAGATATATTTCTATTTAGTGAAGCGGAGAAACA 660
Db	601	GAAAAGCAGAACACACGAGATATATTTCTATTTAGTGAAGCGGAGAAACA 660
QY	661	GGCGAAGAGCTGTCGGTGAAGAGCTGATTCGTTTTCACGCTGCTGCTGCGCGGA 720
Db	661	GGCGAAGAGCTGTCGGTGAAGAGCTGATTCGTTTTCACGCTGCTGCTGCGCGGA 720
QY	721	AATGAACACCTACAACTGATTTCAATGCGATGTACAGATATTAAGAACGCGAGC 780
Db	721	AATGAACACCTACAACTGATTTCAATGCGATGTACAGATATTAAGAACGCGAGC 780
QY	781	GTTTACGAGAACTGCGCAGCCTCTGAACTGATGCTCAGCGAGTGAAGAGCTTTG 840
Db	781	GTTTACGAGAACTGCGCAGCCTCTGAACTGATGCTCAGCGAGTGAAGAGCTTTG 840
QY	841	CGTTTACAGCGCGCGCGCTTTTGGGCGCATTTGCCAAGCGGATACGAGATCGG 900
Db	841	CGTTTACAGCGCGCGCGCTTTTGGGCGCATTTGCCAAGCGGATACGAGATCGG 900
QY	901	GGGACCTGATTAAGAGAGTGAATGTTTGGGCTTTTGGCATTCGGCAATCGTGT 960
Db	901	GGGACCTGATTAAGAGAGTGAATGTTTGGGCTTTTGGCATTCGGCAATCGTGT 960
QY	961	GAAGCAAGTTTGAAGAGAGTGAATGTTTGGGCTTTTGGCATTCGGCAATCGTGT 1020
Db	961	GAAGCAAGTTTGAAGAGAGTGAATGTTTGGGCTTTTGGCATTCGGCAATCGTGT 1020
QY	1021	GCGTTTGGCCACGCGCATCCATTTTTCGCTTGGGCGCGCTTGGCGCTTTGAAGCAAT 1080
Db	1021	GCGTTTGGCCACGCGCATCCATTTTTCGCTTGGGCGCGCTTGGCGCTTTGAAGCAAT 1080
Db	1081	ATCGCGTTAAACGCTTTTGAATTTCTGCTTTTCTCATATGAGTGGTGGTCACTCCG 1140
Db	1081	ATCGCGTTAAACGCTTTTGAATTTCTGCTTTTCTCATATGAGTGGTGGTCACTCCG 1140
QY	1141	ATTGAAAACAGTGTGATATACGGAATTAAGAGCTTCGCTGTAAGATGTAA 1191
Db	1141	ATTGAAAACAGTGTGATATACGGAATTAAGAGCTTCGCTGTAAGATGTAA 1191
RESULT 2		
AAA95667		
ID	AAA95667	standard; DNA; 1204 BP.
XX	AAA95667;	
DT	14-FEB-2001	(first entry)
DE	Bacillus subtilis	hydroxylating enzyme gene #2.
XX	Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;	
KW	Hydroxymethylglutaryl co-enzyme A reductase; hypcholesterolaemia.	
XX	Bacillus subtilis.	
XX	WO200444886-A1.	
PN	03-AUG-2000.	
PD	28-JAN-2000;	2000WO-JP000472.
XX	29-JAN-1999;	99JP-00021707.
PR	(KYOW) KYOWA HAKKO KOGYO KK.	
PA	Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;	
PI	WPI; 2000-548827/50.	
XX	P-PSDB; AAB15502.	
DR	New protein derived from Bacillus genus microorganism useful for	
PT	producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase	
PT	inhibitors.	
XX	Claim 16; Page 90-94; 111pp; Japanese.	
PS	This sequence represents the coding region for a novel Bacillus derived	
CC	protein having the activity of producing a hydroxylated bicyclic compound	
CC	or the corresponding lactone from a bicyclic compound or the	
CC	corresponding lactone. The protein is used for preparing hydroxylated	
CC	compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase	
CC	inhibitors and hypcholesterolaemic agents	
XX	Sequence 1204 BP; 333 A; 253 C; 316 G; 302 T; 0 U; 0 Other;	
SQ	Query Match 99.6%; Score 1186.2; DB 3; Length 1204;	
	Best Local Similarity 99.7%; Pred. No. 0;	
	Matches 1188; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY	1	ATGATGTGTTAAACCGCGGCAAGCTTTGACGAGCGCTGCTCAATGGGAAACAA 60
Db	8	ATGATGTGTTAAACCGCGGCAAGCTTTGACGAGCGCTGCTCAATGGGAAACAA 67
QY	61	CAGGATGCGTATCATCCGTTTCATCGTATGAATCGATGAGAAAGATGCGCTGTTCC 120
Db	68	CAGGATGCGTATCATCCGTTTCATCGTATGAATCGATGAGAAAGATGCGCTGTTCC 127
QY	121	TTTCATGAAGAAACCAAGTGTGAGCGTTTCTTTATGATGATGTCAAAAAGTTGT 180
Db	128	TTTCATGAAGAAACCAAGTGTGAGCGTTTCTTTATGATGATGTCAAAAAGTTGT 187
QY	191	GGGATAAAGAGTTGTTTTCAGTTGCTGCGGACGACAGCTCTATTGGAATTC 240

```
Db 188 GGGGTAAGAGTGTGTTTCCAGTTGTCATGCCGACACACAGCTCTATTGGAAATCC 247
Qy 241 ATCATTAACATGACCGCCGCGAGCATACAAAATCCGTTCACTGCTGAAACAAAGCCTT 300
Db 248 ATCATTAACATGACCGCCGCGAGCATACAAAATCCGTTCACTGCTGAAACAAAGCCTT 307
Qy 301 ACTCCGCGGTGATGAAGCATGGGAACCGGAGATTCAGAAATCAAGATGAACCTGATT 360
Db 308 ACTCCGCGGTGATGAAGCATGGGAACCGGAGATTCAGAAATCAAGATGAACCTGATT 367
Qy 361 CAAAATTTTCAGGCGCGAGTCAGTTGTCACCTTGTTCAGATTTTTCATCCCGTCCG 420
Db 368 CAAAATTTTCAGGCGCGAGTCAGTTGTCACCTTGTTCAGATTTTTCATCCCGTCCG 427
Qy 421 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCATATGAACAGTTTAAAGCA 480
Db 428 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCATATGAACAGTTTAAAGCA 487
Qy 481 TGGTCTGATCTTCTGTCAGTACACCGAAGATTAAGTGAAGAGCTGAAAAGCCTTT 540
Db 488 TGGTCTGATCTTCTGTCAGTACACCGAAGATTAAGTGAAGAGCTGAAAAGCCTTT 547
Qy 541 TTGGAAGAACGAGATAAGTGTGAGAGAACTGGCGCGTCTTTTTCGCGCATATAGAA 600
Db 548 TTGGAAGAACGAGATAAGTGTGAGAGAACTGGCGCGTCTTTTTCGCGCATATAGAA 607
Qy 601 GAAAAGCGAAACAAACCGGAACAGGATATTTTCTATTATTAGTGAAGCGGAAGAAACA 660
Db 608 GAAAAGCGAAACAAACCGGAACAGGATATTTTCTATTATTAGTGAAGCGGAAGAAACA 667
Qy 661 GGGCAGAGCTGTCGCGTGAAGAGCTGATTCGCTTTGACGCTGCTGCTGTCGCGGGA 720
Db 668 GGGCAGAGCTGTCGCGTGAAGAGCTGATTCGCTTTGACGCTGCTGCTGTCGCGGGA 727
Qy 721 AATGAACCACTACAAACCTGATTTCAATCGCATGTACACATATTAGAAACGCCAGGC 780
Db 728 AATGAACCACTACAAACCTGATTTCAATCGCATGTACACATATTAGAAACGCCAGGC 787
Qy 781 GTTTACGAGAACTGCGCAGCATCTGTAAGTCTGCTCAGGAGTGAAGAGCCTTG 840
Db 788 GTTTACGAGAACTGCGCAGCATCTGTAAGTCTGCTCAGGAGTGAAGAGCCTTG 847
Qy 841 CGTTTCAGAGCGCGCCCGCTTTGAGCGCATTCGCAAGCGGGATACGGAGATCGGG 900
Db 848 CGTTTCAGAGCGCGCCCGCTTTGAGCGCATTCGCAAGCGGGATACGGAGATCGGG 907
Qy 901 GGCACCTGATTAAAGAGGTGATATGTTTGGCGTTTGTGGCATCGGCAAAATCGTGAT 960
Db 908 GGCACCTGATTAAAGAGGTGATATGTTTGGCGTTTGTGGCATCGGCAAAATCGTGAT 967
Qy 961 GAAGCAAGTTTGACAGACCGCACTGTTTGAATCCGCGCATCCCAATCCGCATATT 1020
Db 968 GAAGCAAGTTTGACAGACCGCACTGTTTGAATCCGCGCATCCCAATCCGCATATT 1027
Qy 1021 GCGTTTGGCAGCGCATCAATTTTGGCTTTGGGCGCGCTTGGCCGCTTTGAAGCAAT 1080
Db 1028 GCGTTTGGCAGCGCATCAATTTTGGCTTTGGGCGCGCTTGGCCGCTTTGAAGCAAT 1087
Qy 1081 ATCGCGTTAACTGCTTTGATTTCTGTTTCTCATATGAGTGGGTCAGTATCACTCCG 1140
Db 1088 ATCGCGTTAACTGCTTTGATTTCTGTTTCTCATATGAGTGGGTCAGTATCACTCCG 1147
Qy 1141 ATTGAAGAACAGTGTGATATACGAGTTAAGAGCTTCCGTTGTAAGATGTAA 1191
Db 1148 ATTGAAGAACAGTGTGATATACGAGTTAAGAGCTTCCGTTGTAAGATGTAA 1198
```

RESULT 3
AAA95668
ID AAA95668 standard; DNA; 1221 BP.
XX
AC AAA95668;

```
XX 14-FEB-2001 (first entry)  
XX Bacillus subtilis hydroxylating enzyme gene #3.  
XX DE Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;  
XX KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.  
XX OS Bacillus subtilis.  
XX PN W0200044886-A1.  
XX PD 03-AUG-2000.  
XX PF 28-JAN-2000; 2000WO-JP000472.  
XX PR 29-JAN-1999; 39JP-00021707.  
XX PA (KYOW) KYOWA HAKKO KOGYO KK.  
XX PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;  
XX WPI; 2000-548827/50.  
XX CC New protein derived from Bacillus genus microorganism useful for  
XX PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase  
XX PS inhibitors.  
XX PS Claim 16; Page 97-100; 111pp; Japanese.  
XX CC This sequence represents the nucleotide sequence for a novel Bacillus  
XX CC derived protein having the activity of producing a hydroxylated bicyclic  
XX CC compound or the corresponding lactone from a bicyclic compound or the  
XX CC corresponding lactone. The protein is used for preparing hydroxylated  
XX CC compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase  
XX CC inhibitors and hypocholesterolaemic agents  
XX SQ Sequence 1221 BP; 336 A; 257 C; 320 G; 308 T; 0 U; 0 Other;  
  
Query Match 99.2%; Score 1181.4; DB 3; Length 1221;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1185; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Qy 1 ATGAATGTGTAAACCGCGCGCAAGCCTTGACGAGCGCTGCTCAATGGGAACAA 60  
Db 25 ATGAACGTCTGAACCGCGCTCAAGCCTTGACGAGCGCTGCTCAATGGGAACAA 84  
  
Qy 61 CAGGATCGGTATCATCCGTTTCCATGGTATGAATCGATGAGAAAGGATCGCCTGTTCC 120  
Db 85 CAGGATCGGTATCATCCGTTTCCATGGTATGAATCGATGAGAAAGGATCGCCTGTTCC 144  
  
Qy 121 TTTCATGAAGAAACCAAGTGTGAGCGTCTTTCTTATGATGATGTCACAAAAGTTGTT 180  
Db 145 TTTCATGAAGAAACCAAGTGTGAGCGTCTTTCTTATGATGATGTCACAAAAGTTGTT 204  
  
Qy 181 GGGGATTAAGAGTTGTTTTCAGTTGCTATGCCGAGCAGACAAGCTCTATTGGAAATCC 240  
Db 205 GGGGATTAAGAGTTGTTTTCAGTTGCTATGCCGAGCAGACAAGCTCTATTGGAAATCC 264  
  
Qy 241 ATCATTAACATGACCGCCGCGAGCATACAAAATCCGTTCACTGCTGAAACAAAGCCTT 300  
Db 265 ATCATTAACATGACCGCCGCGAGCATACAAAATCCGTTCACTGCTGAAACAAAGCCTT 324  
  
Qy 301 ACTCCGCGGTGATGAAGCATGGGAACCGGAGATTCAGAAATCAAGATGAACCTGATT 360  
Db 325 ACTCCGCGGTGATGAAGCATGGGAACCGGAGATTCAGAAATCAAGATGAACCTGATT 384  
  
Qy 361 CAAAATTTTCAGGCGCGAGTCAGTTGTCACCTTGTTCAGATTTTTCATCCCGCTTCCG 420  
Db 385 CAAAATTTTCAGGCGCGAGTCAGTTGTCACCTTGTTCAGATTTTTCATCCCGCTTCCG 444  
  
Qy 421 GTTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCATATGAACAGTTTAAAGCA 480
```

445 GTTATTGTGATATCTGAGCTGCTGGAGCTGCTTCAGCGCATATGGAACAGTTTAAAGCA 504
481 TGGTCTGATCTTCTGCTCAGTACCGGAGGATATAAGTGAAGAGCTGAAAGAGCCCTTT 540
505 TGGTCTGATCTTCTGCTCAGTACCGGAGGATATAAGTGAAGAGCTGAAAGAGCCCTTT 564
541 TTGGAAGAACAGATAGTGTGAGGAAGAACTGGCCGCGTTTTCGCCGCGCATCATGAA 600
565 TTGGAAGAACAGATAGTGTGAGGAAGAACTGGCCGCGTTTTCGCCGCGCATCATGAA 624
601 GAAAGCGAAACAAACCGGAACAGATATTTCTATTTTAGTGGAAAGCGGAAGAAC 660
625 GAAAGCGAAACAAACCGGAACAGATATTTCTATTTTAGTGGAAAGCGGAAGAAC 684
661 GCGGAGAGCTGCTCGGTGAAGAGCTGATTCGTTTTCAGCGCTGCTCTGTCGCCGGA 720
685 GCGGAGAGCTGCTCGGTGAAGAGCTGATTCGTTTTCAGCGCTGCTCTGTCGCCGGA 744
721 AATGAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTAAGAACGCCAGC 780
745 AATGAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTAAGAACGCCAGC 804
781 GTTACGAGGAAGCTCGGAGCGATCTGAACTGATGCTCAGGCGAGTGSAGAGCCCTTG 840
805 GTTACGAGGAAGCTCGGAGCGATCTGAACTGATGCTCAGGCGAGTGSAGAGCCCTTG 864
841 CGTTTCAGAGCGCGGCCCGCTTTTGGGCGCATTTGCCAAGCGGATACCGAGATCGG 900
865 CGTTTCAGAGCGCGGCCCGCTTTTGGGCGCATTTGCCAAGCGGATACCGAGATCGG 924
901 GGGCACCTGATTAAGAGAGTGTATGCTTTTGGCGTTTGTGGCATCGGCAAAATCGTGAT 960
925 GGGCACCTGATTAAGAGAGTGTATGCTTTTGGCGTTTGTGGCATCGGCAAAATCGTGAT 984
961 GAACCAAGTTTGAAGACCGCATGTTTATATCCGCGCGCATCCCAATCCGCAATTT 1020
985 GAACCAAGTTTGAAGACCGCATGTTTATATCCGCGCGCATCCCAATCCGCAATTT 1044
1021 GCGTTTGGCCACGGCATCCATTTTGGCTTGGCGCGCGCTTGGCGCTTTTGAAGCAAT 1080
1045 GCGTTTGGCCACGGCATCCATTTTGGCTTGGCGCGCGCTTGGCGCTTTTGAAGCAAT 1104
1081 ATCCGTTTAAGCTTTTGATTTCTGCTTTTCTCATATGAGAGTGGTCAATCACTCCG 1140
1105 ATCCGTTTAAGCTTTTGATTTCTGCTTTTCTCATATGAGAGTGGTCAATCACTCCG 1164
1141 ATTGAACACAGTGTATATACGGATTAAGAGCTTCGCTGTGAAATGTA 1191
1165 ATTGAACACAGTGTATATACGGATTAAGAGCTTCGCTGTGAAATGTA 1215

RESULT 4

AAA95669

ID AAA95669 standard; DNA; 1221 BP.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX (KYOW) KYOWA HAKKO KOGYO KK.
XX PA
XX Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX WPI; 2000-548827/50.
XX P-PSDB; AAB15503.
XX
XX New protein derived from Bacillus genus microorganism useful for
XX producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
XX inhibitors.
XX
XX Claim 16; Page 101-104; ilipp; Japanese.
XX
XX This sequence represents the coding region for a novel Bacillus derived
XX protein having the activity of producing a hydroxylated bicyclic compound
XX or the corresponding lactone from a bicyclic compound or the
XX corresponding lactone. The protein is used for preparing hydroxylated
XX compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
XX inhibitors and hypocholesterolaemic agents
XX
XX Sequence 1221 BP; 334 A; 261 C; 322 G; 304 T; 0 U; 0 Other;

Query Match 98.1%; Score 1168.6; DB 3; Length 1221;

Best Local Similarity 98.8%; Pred. NO. 0;

Matches 1177; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 ATGATGTGTTAAACCGCGCGCAAGCCCTTCAGCGAGCGCTCAATGGGAAAAACAAA 60
DB 25 ATGACGTTCTGACCGCGCTCAAGCCCTTCAGCGAGCGCTCAATGGGAAAAACAAA 84
QY 61 CAGGATGCGTATCATCCGTTTCCATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 85 CAGGATGCGTATCATCCGTTTCCATGATGATGATGATGATGATGATGATGATGATGAT 144
QY 121 TTGATGAAGAAACCAAGTGTGGAGCGTTTCTTATGATGATGATGATGATGATGATGAT 180
DB 145 TTGATGAAGAAACCAAGTGTGGAGCGTTTCTTATGATGATGATGATGATGATGATGAT 204
QY 181 GGGGATAAAGAGTTGTTTTCAGTTGCTGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 240
DB 205 GGGGATAAAGAGTTGTTTTCAGTTGCTGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 264
QY 241 ATCATTAACATGAGCCCGCGGAGCATACAAATATCCGTTGATGATGATGATGATGATGAT 300
DB 265 ATCATTAACATGAGCCCGCGGAGCATACAAATATCCGTTGATGATGATGATGATGATGAT 324
QY 301 ACTCGCGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 325 ACTCGCGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
QY 361 CAAAAATTTTCAGGGCGCAGTGAAGTTTTCAGTTGCTGCGCAGCAGCAGCAGCAGCAGCAGCAG 420
DB 385 CAAAAATTTTCAGGGCGCAGTGAAGTTTTCAGTTGCTGCGCAGCAGCAGCAGCAGCAGCAGCAG 444
QY 421 GTTATTGTGATATCTGATGCTGCTGAGTGTGCTTTCAGCGCAGTGAAGAGTGAAGAGTGAAG 480
DB 445 GTTATTGTGATATCTGATGCTGCTGAGTGTGCTTTCAGCGCAGTGAAGAGTGAAGAGTGAAG 504
QY 481 TGGTCTGATCTTCTGCTCAGTACACCGGAGGATATAAGTGAAGAGCTGAAAGAGCCCTTT 540
DB 505 TGGTCTGATCTTCTGCTCAGTACACCGGAGGATATAAGTGAAGAGCTGAAAGAGCCCTTT 564
QY 541 TTGGAAGAACAGATAGTGTGAGGAAGAACTGGCCGCGTTTTCGCCGCGCATCATGAA 600
DB 565 TTGGAAGAACAGATAGTGTGAGGAAGAACTGGCCGCGTTTTCGCCGCGCATCATGAA 624
QY 601 GAAAGCGAAACAAACCGGAACAGATATTTCTATTTTAGTGGAAAGCGGAAGAAC 660
DB 625 GAAAGCGAAACAAACCGGAACAGATATTTCTATTTTAGTGGAAAGCGGAAGAAC 684
QY 661 GCGGAGAGCTGCTCGGTGAAGAGTGTGATTCGTTTTCAGCGCTGCTCTGTCGCCGGA 720

Db 926 AAGGGTTCTACAGTGATGAGTGGGTGGCTTCAGCGAATCGTGACGAATTAAGTTGA 985
 QY 975 CAGACCGCACATGTTGATATCCGCGCCCATCCCAATCCGCATATTGCGTTTGCCACGG 1034
 Db 986 CGATCCTGACAGCTTCAAGCTTGATCGCAATCAAACTTCATAGAGCTTCGGCTCGG 1045
 QY 1035 CATCCATTTTGGCTTGGGGCCCGCTTGGCCGCTTGAAGCAAAATATCGGTTAAGTC 1094
 Db 1046 CATCCATTTTGGCTTGGGGCCCGCTTGGCCGCTTGAAGCAAAATATCGGTTAAGTC 1105
 QY 1095 TTGAT 1100
 Db 1106 CTGCT 1111

RESULT 7

AAD5813
 ID AAD5813 standard; DNA; 1233 BP.

XX AAD5813;

XX 27-OCT-2003 (revised)

DT 07-AUG-2003 (first entry)

DE Micromonospora carbonacea polyketide synthase (PKS) type I gene #3.

KW Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme; gene; ds.

OS Micromonospora carbonacea.

XX Key Location/Qualifiers

EH 1.1233

FT CDS

FT /*tag= a

FT /product= "Polyketide synthase"

XX CA2391131-A1.

XX 19-NOV-2002.

XX 26-JUL-2002; 2002CA-02391131.

XX 26-JUL-2001; 2001US-0307629P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Yang X, Staffa A, Farnet CW;

XX WPI; 2003-343556/33.

XX P-PSDB; AAE36997.

XX Novel isolated polypeptide involved in biosynthesis of macrolides by

XX microorganisms, useful for biosynthesis of macrolides by microorganisms,

XX preferably for biosynthesis of rosaramicin.

XX Claim 1; Page 101-102; 206pp; English.

XX The invention relates to genes and proteins involved in the biosynthesis

XX of macrolides by microorganisms. In particular it relates to the nucleic

XX acids forming the biosynthetic locus for rosaramicin (a 16-member

XX macrolide antibiotic) from Micromonospora carbonacea. The invention is

XX useful for the biosynthesis of macrolides by microorganisms. It allows

XX direct manipulation of macrolides and related chemical structures by

XX chemical engineering of the proteins involved in the biosynthesis of

XX rosaramicin. It is useful to catalyse certain biochemical reactions, in

XX vitro or in vivo, to direct or enhance the synthesis or modification of a

XX polyketide, polyketide substrate or its precursor. The present sequence

XX is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-

XX OCT-2003 to standardise OS field)

XX Sequence 1233 BP; 225 A; 452 C; 377 G; 179 T; 0 U; 0 Other;

XX Query Match 12.5%; Score 149; DB 7; Length 1233;

Best Local Similarity 48.5%; Pred. No. 4.2e-37;
 Matches 410; Conservative 0; Mismatches 435; Indels 0; Gaps 0;
 QY 239 CCATCAATTAACATGAGACCCCGCGAAGCATACAAATAATCCGTTAGTCGTGAAACAAAGCCT 298
 Db 254 CCATCGCGGCATCGACCCCGCGGCGACCGCGCGCTGCGCAAGCTGTGAGCCAGCGGT 313
 QY 299 TTACTCCGCGGTGATGAAGCAATGGGAACCGAGAATTCAGAAATCACAGATGAATCTGA 358
 Db 314 TCACCCCGCGCGATCGCCAGATGAGCTGGGATCGGGCAGATCACCGCCAGCGTGC 373
 QY 359 TTCAAAATTTTACGGGGCGAGTTCAGTTTCACCTTGTTCAGATTTTTCATACCCGCTTC 418
 Db 374 TCAGACAGGTACCGACCGAGACCGGATCGACATCGCCAGCGCTCGGTACCGCTGC 433
 QY 419 CGGTATTGTATATCTGAGCTCTGGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAG 478
 Db 434 CGGTACGGTCNTCCGCGAGCTGCTGGCATTCCTCCACCAAGGATCAGAGNAGTTCCGCG 493
 QY 479 CATGCTCTGATCTTCTGCTCAGTACACCGAAGGATAAAGTGAAGAAGCTGAAAAGCCT 538
 Db 494 AGTGGGTGGACATCATCTCAGCAACGAAGGGCTGGAGTATCCCAAGCTCCCGGACGACT 553
 QY 539 TTTTGGAGACCGAGATAAGTGTGAGGAAGTCTGGCGGCTTTTTCGCGGATCTAG 598
 Db 554 TCACCGGAGCGTGGGCGCCCGCCATCGAGGAGTGTCCGAATCTCTGTACGCCAGATCG 613
 QY 599 AAGAAAACGGAACAAACCCGGAACAGGATATTATTTCTATTATTAGTGAACCGGAAGAAA 658
 Db 614 CCCACAGCGCGCGAACCAGAACCGACCTGTATCAGCGCTCTGTGCGCGGAGGTGCG 673
 QY 659 CAGCGAGAGCTGTCCGGTGAAGAGCTGATTCGCTTTTGACCGCTGTCTGTGTGGCGG 718
 Db 674 ACGGCGCAAGCTGACCGGACGAGAGTGTCTCAACATCGTCGCGCTCTCTGCTCAACGCG 733
 QY 719 GAATGAACCACTACAAACCTGATTTCAATGGATGTACAGCATATTAGAAACGGCAG 778
 Db 734 GGCACATCTCCAGCGCACGCTGCTCAGCAACCTGTTCTCTGGTGTCTGGAGGAGCACCGCG 793
 QY 779 GCGTTTACAGGAACCTCGCAGCGCATCTGAACTGATGCTTCAGGCGATGAGGAAGCCT 838
 Db 794 AGGCACAGCGCGGTTCGCGCGCGACCGCAGCTCTGTGCGGGGTGNTCCAGGAGCGC 853
 QY 839 TGGTTTTCAGAGCGCGCGCGCGTGTGAGGGGCATTTGCCAAGCGGAGTACGAGATCG 898
 Db 854 TGGCTACCGGTCCCGCTTCAACTGCACTCTTCGGAATCTTGAACGAGACACCGACATCC 913
 QY 899 GGGGCGACCTGATTAAGAAAGTGTATATGTTTGGGCTTTGTGCGCATCGGCAATCTGT 958
 Db 914 TCGGCCACCCCATCGCGCAAGGGCCAGATGTGTATCGGCTGGATCGCTTCCGCGAACCGCG 973
 QY 959 ATGAAGCAAAAGTTTGACAGACCGCACATGTTTGATATCCGCGCGCATCCCAATCCGATA 1018
 Db 974 ACACCGAGGTGTTTACCGACCGCGACACTTTCGACATCCGACCGAGTTCGACCAAGCACC 1033
 QY 1019 TTGCGTTTGGCCACCGCATTCATTTTTCCTTGGGGCCCCCGCTTGCCTTCTTGAAGCAA 1078
 Db 1034 TGGCGTTTGGGCCACCGCATCCACCACTGCTGGGCGGGTTCCTGGCCAGGCTGGAGGCGA 1093
 QY 1079 ATATC 1083
 Db 1094 AGGTC 1098

RESULT 8

AAD5810/c

ID AAD5810 standard; DNA; 60196 BP.

XX AAD5810;

XX AC

XX 27-OCT-2003 (revised)

DT 07-AUG-2003 (first entry)

XX

DE Micromonospora carbonacea polyketide synthase (PKS) type I gene cluster.

XX Macrolide; rosaramycin; polyketide; polyketide synthase; PKS; enzyme;
KW gene; ds.

XX Micromonospora carbonacea.

XX Key Location/Qualifiers

FT CDS 1..1683
FT /tag= a
FT /product= "Polyketide synthase #1"
FT /note= "CDS does not include start codon"

FT /partial
FT complement(1728..2522)

FT CDS /tag= b

FT /product= "Polyketide synthase #2"

FT complement(2629..3861)

FT CDS /tag= c

FT /product= "Polyketide synthase #3"

FT 4365..5573

FT /tag= d

FT /product= "Polyketide synthase #4"

FT 5702..19117

FT CDS /tag= e

FT /product= "Polyketide synthase #5"

FT 19144..24921

FT /tag= f

FT /product= "Polyketide synthase #6"

FT /note= "CDS does not include start codon"

FT /partial

FT 24993..36230

FT CDS /tag= g

FT /product= "Polyketide synthase #7"

FT 36292..41016

FT CDS /tag= h

FT /product= "Polyketide synthase #8"

FT 41049..46403

FT CDS /tag= i

FT /product= "Polyketide synthase #9"

FT 46400..47794

FT CDS /tag= j

FT /product= "Polyketide synthase #10"

FT /note= "CDS does not include start codon"

FT /partial

FT 47794..49083

FT CDS /tag= k

FT /product= "Polyketide synthase #11"

FT 49092..49814

FT CDS /tag= l

FT /product= "Polyketide synthase #12"

FT 49868..51226

FT CDS /tag= m

FT /product= "Polyketide synthase #13"

FT /transl_except= (pos:51221..51223, aa:Xaa)

FT /note= "Xaa corresponds to amino acids from position 452-1811"

FT 51506..53416

FT CDS /tag= n

FT /product= "Polyketide synthase #14"

FT /note= "CDS does not include start codon"

FT /partial

FT complement(54569..53358)

FT CDS /tag= o

FT /product= "Polyketide synthase #15"

FT 54897..56342

FT CDS /tag= p

FT /product= "Polyketide synthase #16"

FT 56408..57634

FT CDS /tag= q

FT /product= "Polyketide synthase #17"

FT 57657..59123

FT CDS /tag= r

FT /product= "Polyketide synthase #18"

FT /note= "CDS does not include start codon"

FT /partial

FT CDS 59363..60196

FT /tag= s

FT /product= "Polyketide synthase #19"

XX CA2391131-Al.

PN 19-NOV-2002.

XX 26-JUL-2002; 2002CA-02391131.

XX 26-JUL-2001; 2001US-0307629P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Yang X, Staffa A, Farnet CM;

XX WPI: 2003-343556/33.

XX P-PSDB; AAE36995, AAE36996, AAE36997, AAE36998, AAE36999, AAE37000,

XX AAE37001, AAE37002, AAE37003, AAE37004, AAE37005, AAE37006, AAE37007,

XX AAE37008, AAE37009, AAE37010, AAE37011, AAE37012, AAE37013.

XX Novel isolated polypeptide involved in biosynthesis of macrolides by

XX microorganisms, useful for biosynthesis of macrolides by microorganisms,

XX preferably for biosynthesis of rosaramicin.

XX Example 2; Page 59-94; 206pp; English.

XX The invention relates to genes and proteins involved in the biosynthesis

XX of macrolides by microorganisms. In particular it relates to the nucleic

XX acids forming the biosynthetic locus for rosaramicin (a 16-member

XX macrolide antibiotic) from Micromonospora carbonacea. The invention is

XX useful for the biosynthesis of macrolides by microorganisms. It allows

XX direct manipulation of the proteins involved in the biosynthesis of

XX chemical engineering of the proteins involved in the biosynthesis of

XX rosaramicin. It is useful to catalyse certain biochemical reactions, in

XX vitro or in vivo, to direct or enhance the synthesis or modification of a

XX polypeptide, polypeptide substrate or its precursor. The present sequence

XX is M. carbonacea polyketide synthase (PKS) type I gene cluster. (Updated

XX on 27-OCT-2003 to standardise OS field)

XX SQ Sequence 60196 BP; 5934 A; 20042 C; 23822 G; 9398 T; 0 U; 0 Other;

Query Match 12.5%; Score 149; DS 7; Length 60196;

Best Local Similarity 48.5%; Pred. No. 4e-36;

Matches 410; Conservative 0; Mismatches 435; Indels 0; Gaps 0;

QY 239 CCATCTTAACATGACCGCGGAGCATACAAAATCGTTCAGTCGTGACAAAGCCT 298

DB 3608 CCATCGCGGCATGACCGCGGAGCATACAAAATCGTTCAGTCGTGACAAAGCCT 3549

QY 299 TTACTCCCGCGGTGATGAAGCAATGGGAACCGAGAAATTCAGAAATCAAGATGAACCTGA 358

DB 3548 TCACCCCGCGCGGTGATGAAGCAATGGGAACCGAGAAATTCAGAAATCAAGATGAACCTGA 3489

QY 359 TCAAAAATTTACGGGGCGGAGTGTGAGTTTTCAGATTTTTCATACCGCTTC 418

DB 3488 TCGACACAGGTACGGGACCGAGGACCGGATCGACATCGCCAGGACCTCGCGTACCGCTGC 3429

QY 419 CGGTATTGTGATATCTGAGCTGCTGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAG 478

DB 3428 CGGTGACGGTCAATCGCGAGCTGCTGGGCAATCCCAACCAAGATCACGAGAAGTTCCGCG 3369

QY 479 CATGCTCTGATCTTCTGGTCACTACACCGAAGGATAAAGTGAAGAGCTGAAAAGCCT 538

DB 3368 AGTGGGTGACATCATCTCTGACCAACGAGGCTGGAGTATCCCAACCTCCCGGACGACT 3309

QY 539 TTTTGAAGAACGAGATTAAGTGTGAGGAGAACTGCGCGCTTTTTCGCGGATCATAG 598

DB 3308 TCACCGACAGCGTGGCGCCGCCCATCGAGGAGTGGTCCGAATTCCTGTACGCCAGATCG 3249

QY 599 AAGAAAGCGAAACAAACCGGAAACAGGATATATTTCTATTATTAGTGAAGCGGAGAA 658


```

Db 3248 CCACAAAGCGCGCGAACCAGACACACCTGATCAGCGCCTCTGTGCGCGGAGGTGG 3189
Qy 659 CAGGCGAGAGCTGCGGTGAAGAGCTGATTCGGTTTTCACGCTGCTGCTGTGCGCG 718
Db 3188 ACGGCGAGAGCTGACCGAGAGAGTCTGACATCTGTCGCGCTGCTGCTACCGCGCG 3129
Qy 719 GAATGAACACCTACAAACCTGATTTCAATCGATGATGATGATGATGATGATGATG 778
Db 3128 GGCACATCTCCAGCGCCAGCTGCTCAGCAACCTGTTCTCTGCTGCTGAGGAGCA 3069
Qy 779 GCGTTTACGAGGAAGCTGCGGAGCACTCTCACTGATGCTCTCAGGAGTGGAGAGC 838
Db 3068 AGGCACAGGCGCGGCTGCGGCGGAGCGAGCTGCTGCGGCGGCTGATCAGGAGAG 3009
Qy 839 TCGGTTTACAGAGCGCGCGCGCGCTGTTTGAGGCGCATGCAAGCGGATACGAGATCG 898
Db 3008 TCGCTACCGGTCGCCGTTTCACTGCACTCTCGGATCTCGGATCTCGGATCTCGG 2949
Qy 899 GCGGCGACCTGATTAAGAAAGGTGATATGTTTGGCGTTTGGCATCGGCAAAATCG 958
Db 2948 TCGGCGACCCCATGCGCAAGGCGAGATGTGATCGCTGATCGCTTCCGCAACCGCG 2889
Qy 959 ATGAAGCAAGTTTGACAGACCGCACTGTTTATATCCCGCGCATCCCAATCCGCA 1018
Db 2888 ACACCGAGGTGTCACGAGACCGGACACCTTCGACATCCGACCGAGTCAACAGCA 2829
Qy 1019 TTGCGTTTGGCCACCGCATCATTTTGGCTTGGGCGCGCGCTGCGCGCTTGAAGCA 1078
Db 2828 TGGCGTTCGCCACCGCATCCACCACTGCTGCGGCGGTTCTGCGCAGGCTGGAGCG 2769
Qy 1079 ATATC 1083
Db 2768 AGGTC 2764

```

RESULT 9
AAQ81792
ID AAQ81792 standard; DNA; 8478 BP.

XX AAQ81792;

XX 25-MAR-2003 (revised)

XX 26-SEP-1995 (first entry)

XX B. subtilis biotin operon and flanking sequences.

XX Biotin operon; recombinant production; dietary additive; animal feed;
KW vitamin supplement; research reagent; ds.

XX Bacillus subtilis.

XX Key Location/Qualifiers

EH Terminator

FT /tag= k

FT /note= "rho-independent termination site"

FT /tag= a

FT /note= "sigma A promoter"

FT /tag= b

FT /note= "potential regulatory site"

FT /tag= c

FT /product= "bioW"

FT /tag= d

FT /product= "bioA"

FT /tag= e

FT /product= "bioF"

FT /tag= f

```

FT CDS /product= "bioD"
FT 4408..5415
FT /*tag= g
FT /product= "bioB"
FT 5423..5462
FT /*tag= i
FT /note= "rho-independent termination site"
FT 5484..6671
FT /*tag= h
FT /product= "bioI"
FT 6748..7509
FT /*tag= i
FT /label= ORF 2
FT 7501..7543
FT /*tag= m
FT /note= "rho-independent termination site"
FT 7695
FT /*tag= j
FT /label= ORF 3

```

FT EP635572-A2.

PN 25-JAN-1995.

XX 13-JUN-1994; 94EP-00108998.

XX 25-JUN-1993; 93US-00084709.

PR 06-MAY-1994; 94US-00239430.

XX (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX Bower SG, Perkins JB, Pero JG, Yocum RR;

XX WPI; 1995-053684/08.

XX Biotin genes, and constructs derived from *Bacillus subtilis* - for

XX improved production of recombinant biotin or biotin precursor for use in

XX e.g. dietary supplements.

XX Example II; Fig 14; 75pp; English.

XX AAQ81792 is the B. subtilis biotin operon and flanking sequences, as part
CC of an expression vector it can be used in the recombinant production of
CC biotin (or biotin precursor protein). The biotin can be used as a dietary
CC additive in animal feeds, and as a vitamin supplement for human
CC consumption. Biotin is also useful as a reagent for research, and
CC diagnostic procedures. (Updated on 25-MAR-2003 to correct PN field.)

XX Query Match 11.7%; Score 139.2; DB 2; Length 8478;
SQ Best Local Similarity 50.4%; Pred. No. 2e-33;
Matches 426; Conservative 0; Mismatches 408; Indels 12; Gaps 3;

Qy 252 GACCCCGCGAGCATACAAAATCCGTCAGTCGTGAACAAAGCCTTATCTCGCGGT 311
Db 5735 GAACCGCCTGATCATAGACGATTGCGGCGCTTGCAGCGGAGCGTTTACGCGAGAAC 5794

Qy 312 GATGAAGCAATGGGAACCGAGATTCAAGAAATCACAGATGAACATTCAAAATTCA 371

Db 5795 GACAGAGATTTCAGCGGTATATCATTTGAACCTGTCATTTGTTGATCAAGTGCA 5854

Qy 372 GGGGCGCAGTGGTTGACCTTGTTCAGATTTTTCATACCCCGCTTCGGTTATGTGAT 431

Db 5855 AGGTAAAAAAGATGAGGTCAATTCGGACTTTCCTTTTAGCAAGTTTGTTCAT 5914

Qy 432 ATCTGAGCTGCTGGAGTGCCTTCAGCGGAGATGAACAGTTTAAAGCATGGTCTGATCT 491

Db 5915 AGCTACATTATAGTGTACCGGAGGAAGATAGGAGCAATTAAGGAGTGGGTGCGAG 5974

Qy 492 TCTGGTCACTACACCGAGGATAAAGTGAAGAGCTGAAAAGCCTTTTGGAGAACG 551

Db 5975 TCTCAT-----TCAACGATTGATTTTACCCGCTCAAGAAAGCATTAACAGAGGCA 6027

Qy	552	AGATAAGTGTGAGGAAACTGGCCGCGTTTTTTGCGGCGATCATGAGAAAGCGAAA	611
Db	6028	ATATTATGGCTGTGCAGCTATG--GCATATTTCAAGAGCTGATTCAAAGAGAAAACG	6085
Qy	612	CAAAACCGAACAGATATTATTTCTATTTTAGTGSAGCGGAAGAACAGCGCAGAGCT	671
Db	6086	CCACCTCTCAACAGATATGATCAGCTGCTCTTGAAGGGAGAGAA--AGGTATGCT	6142
Qy	672	GTCCGSGTAAGAGCTGATTCGCTTTTGACGCTGCTGCTGGCGCGGAAATGAAACCAAC	731
Db	6143	GACGGAAGAGAGCGCGCATCTACGTGCATATTGCTGGCGATCGCCGACATGAGACAC	6202
Qy	732	TACAAACCTGATTTCAATCGGATGTACAGCATATTAGAAACGCCAGCGGTTTACGAGGA	791
Db	6203	GGTCAATCTCATCAGCAATTCAGTCCTTTGTCTGCTGCAGATCCAGAACAGCTTTTGAA	6362
Qy	792	ACTGCGCAGCATCTCTGAACTGATGCTCTCAGGCAGTGGAGGAAGCCTTGCGTTTCAGAGC	851
Db	6263	ACTGAGAGAAATCCAGATCTTATTGGTACCGCAGTCGAGGAATGTTTACGCTATGAAG	6322
Qy	852	GCGCGCCCCGGTTTTGAGGGCGATTTGCCAAGCGGATACGAGATCGGGGGGACCTGAT	911
Db	6323	CCCCACGCAATGACACGCCAGAGTTGCGTCAGAGGATATTGACATCTCGCGGGTGACGAT	6382
Qy	912	TAAAGAGGTGATATGTTTTTGCGTTTGTGGCATCGGCAATCGTGATGAAGCAAAAGTT	971
Db	6383	CGGTCAAGGAGAACAGTCTATCTTTGTTAGAGCGGCTAATCGAGACCTAGCATATT	6442
Qy	972	TGACAGACCGCACATGTTTGATATCGCCGCGCATCCCAATCGGCATATCGGTTTGCGCA	1031
Db	6443	CACGAACCCCGATGTCTTCGATATTACGAGAAAGTCTTAATCCGCATCTTTCAITTCGGGCA	6502
Qy	1032	CGGCATCCATTTTGGCTTCGSSGCCCGCTTGCCCGTCTTGAAGCAAAATATCGCGTTTAAAC	1091
Db	6503	TGCGCATCATGTTTGCTTTAGGGTCTCTCGCTGGCACGATTAGAGCGGCAATTCGGATTA	6562
Qy	1092	GTCTTT	1097
Db	6563	CACTCT	6568

RESULT 10
ACF06128
ID ACF06128 standard; DNA; 1215 BP.

```

RESULT 10
ACF06128
ID ACF06128 standard; DNA; 1215 BP.
XX
XX ACF06128;
AC
XX
XX DT 30-SEP-2003 (first entry)
XX
XX DE Bacterial P450 enzyme encoding DNA SEQ ID NO:39;

```

XX Bacterial; p450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
 KW diol; alkene; chirality; thermotolerance; thermostability; gene; ds.
 XX Eubacteria.
 CS

AA	Key	Location/Qualifiers
PH	CDS	1..1215
FT		/*tag= a
FT		/product= "p450 enzyme"

XX PN WO2003052050-A2.

26-JUN-2003.

AA
PF 05-AUG-2002; 2002WO-US024910.

PR 03-AUG-2001: 2001US-0309497P.

XX
PA (DIVE-) DIVERSA CORP.XX
PT
Weiner D. Burke M. Hitchman

XX	WFI; 2003-541641/51.
DR	P-PSDB; ABR82138.
DR	
XX	
XX	Novel polypeptide having P450 enzyme activity and polynucleotides
PT	encoding the polypeptide, useful for catalyzing the hydrolysis of
PT	epoxides and arene oxides to their corresponding diols.
XX	
XX	Claim 18; Page 72-73; 365pp; English.
PS	
XX	
XX	The present sequence encodes a bacterial P450 enzyme. P450 enzymes can be
CC	used to catalyse the hydrolysis of epoxides and arene oxides to their
CC	corresponding diols. P450 enzymes can also be used for hydrolysing an
CC	alkene, for producing a compound of a desired chirality, and for
CC	increasing thermotolerance or thermostability of a P450 polypeptide
XX	
XX	Sequence 1215 BP; 168 A; 414 C; 474 G; 159 T; 0 U; 0 Other
SQ	

Query Match	10.1%	Score 120.4;	DB 8;	Length 1215;
Best Local Similarity	49.6%	Pred. No. 8.2e-28;		
Matches 418;	Conservative	0;	Mismatches 391;	Indels 333;
				Gaps 3;

241	Qy	ATCATTTAA	CATGACCGCCGGAAGCATACAA	AAATCGGTT	CAGTCGTGACAAAGCC	TTTT	300	
277	Db	ATCCTCAGCATGACCGCGCGGACCA	CACCGGTTGCGCACCC	TGTCGTCGTCGCGGAGGCG	GTTC	336		
301	Qy	ACTCCGGCGGTGATGAGCAATCGG	AACCGAGAAATTC	CAAGAAATCA	CAGATGAACTGAT	360		
337	Db	ACCATGACACAGGTGAGAAAGTT	CGGCCCGCGGTGCGGAGCT	TGSCCGACGAGCT	GATC	396		
361	Qy	CAAAAAATTT	CAGGGCGCAGTGA	---GTTTGAC	CTTTGTTT	CACGATTTTTCATACCCCGCTT	417	
397	Db	GACAGATG	TGCCCA	CCGGCGCCCGGT	CGACCTG	TCGAGGAGTTCGCGCTCCCGTG	456	
418	Qy	CCGGTTATT	TGTGATATCTGAGCTG	CTGGAGTGC	CTT	CAGCGAGATG	GAACAGTTTTAA	477
457	Db	CCGGTCGGGGTGATCTGCCAGCTG	CTCGCGGTG	CCCGGTTCGAGGACCG	TC	CGCGTTC	CGG	516
478	Qy	GCATGGTCTGATCTTCTGGTTCAGT	ACACCGAAGAT	AAAAAGT	GAAGAGCT	GA	AAAGCC	537
517	Db	CGCTGGAGCGACCGCGCGCTGTCC	ACAGTTCCTG	CACGGCCGAGGAGTT	TCGACG	CGCAAC	576	
538	Qy	TTTTTGGAAACGAGATAGTGTG	AGGAGAACTG	GC	CCCGGTTTTTTT	TCGCGGCA	CATA	597
577	Db	C-----	-----	AGGAGAACTGCGGCGCTT	ATATG	CGGGGGTT	GATC	612

Qy	598	GAGAAAGCGGAACAAACCGGAACAGGATATTATTTCTATTTTAGTGGGAACCGGAAGAA	657
Db	613	GAGGATCACCGGCGGCGTCCGCGTGAGGACCTGATCAACCGGGGTGATCGAGGGCCCGGGAC	672
Qy	658	ACAGGCGGAAGACTGTCCGCTGAAGAGCTGATTCCTGTTTTCGACGCTGCTGCTGGTGCC	717
Db	673	CGGACGACCGGCTGACCGGACGAGAGTTGGTGACCTGTGCTGGTCGATCTCTGGTGCC	732
Qy	718	GGAAATGAAACCACTACAAACCTGATTTCAAAATCCGATGTACAGCATATTAGAAACGCCA	777
Db	733	GGCCACGAGACCAACCGCACGAGATCCCCAACTTCGTGGTGAAGCTGTCTGGACCCGGCCC	792
Qy	778	GGCGTTTACGAGGAACTGCGGACGCATCTCTGAAGTATGCCTCAGGCAGTGGAGGAAGCC	837
Db	793	GAGCAGTGGAAACCGGCTGCGGGAGAGACCGGAGCTGTCTCCGACCGCGGTCTGAGAGACTG	852
Qy	838	TTGCGTTTCAGACGCGCGGCCCGCGTT-----TTGAGGGCGCATTCGCAAGCGGATACG	891
Db	853	ATGCGTTTCGTCCGCTTGGGACGGGTGCCTCGTTTCCCGCGTATCGCCACCGAGGACGTG	912
Qy	892	GAGATCGGGGGCACCTGTATTAAAGAAAGGTGATATGTTTTCGCGTTTGTGGCATCGGCA	951
Db	913	GAGGTGGCGGCACGCTGGTTCGCGCCGGGGACCGGTGCTGTGTGGCGCTGGGGCGGCC	972
Qy	952	AATCGTGATGAAGCAAGATTGTGACAGACCGCACATGTTTGATTTCCGCGCGCCATCCCAAT	1011
Db	973	AACCGCACCGCGCCAGTTTCGACGCGCCCGCAGGAGCTGGACCTCTGGCCCGGAGGGCAAC	1032

```
QY 1012 CCGCATATTGGCTTTGGCCACCGCATCCCATTTTGGCTTTGGGCCCCCGCTTTCCTT 1071
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1033 CAGCAGCTCGGGTTCGGCCATGGCGTCCACCACTGCTCTGGGGGGCGCGCTG 1092
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1072 GA 1073
DB ||
QY 1093 GA 1094
DB ||

RESULT 11
AAI99682_39/c
Continuation (40 of 45) of AAI99682 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2700001
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Query Match 9.9%; Score 118.2; DB 4; Length 110000;
Best Local Similarity 48.1%; Pred. No. 5.9e-26;
Matches 401; Conservative 0; Mismatches 423; Indels 3; Gaps 2;

QY 241 ATCAATTAACATGACCCCGCGAGCATACAAAATCCGTTTCAGTCGTGAACAAGCCCTT 300
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 55295 ATGATCGATAGGCAATCCCGACATCTGTTCCGGCGCAAGCTGTTAAGCGCGCTTC 55236
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 ACTCCCGCGCTGATGAAGCAATGGGACCGAGAAATTCAGAAATCAAGATGAATGATT 360
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 55235 ACCCGAAGCGGTGAAGGACAGGAGCGCTGATTCGCCGCTGTGTGACACCTGATC 55176
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 361 CAAAATTTTGGGGCGCAGTGAGTTTGACCTTGTTTCAAGATTTTTCATACCCGCTCCG 420
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 55175 GACGCGTGTGGAAACGCGCGAGTGAGCTTCTGTCGGGACCTTGGCCGCGCGCTACCG 55116
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 GTTATTGTGATATCTGAGCTGCTGGGAGTGCCCTTTCAGCGCAGATGGAACATGTTTAAAGCA 480
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 55115 ATGGCGGTGATCGGCGACATGCTCGGGTGGCTCCAGAGCAGCGGACATGTTCTTGGCG 55056
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 TGGTCTGATCTTCTGCTCAGTACACCGAAGGATAAAGTGAAGAGCTGAAAAGCTTTT 540
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 55055 TGGTCCGAGCATCTGGTGA-CATTCTCAGTTCGCAATGTTCTCAAGAGGATTTCCAGAT 54997
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 TTGGRAAGCAGATAAAGTGTGAGGAAGACTGCGCGCTTTTTCGCGCATCATAGAA 600
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54996 CACCATGGACGCTTCGCGCCTTACACGACTTCACCGG-----GCCACCATTGGC 54945
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 GAAACGGAACAAACCCGGAACAGGATATTATTCTATTATTAGTGAAGCGGAAGAAACA 660
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54944 GCACGCGAGCGGACCCACCGACGACCTGGTCAGCGTGTCTGGTGAAGTTGAC 54885
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 661 GCGGAGAAGCTGTCGGTGAAGAGCTGATTCGTTTGGACGCTGCTGCTGGTGGCGGA 720
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54884 GCGAGCGGCTAAGCGACGACGAGCTGGTCAAGGAGACGCTGCTGATCCTGATCGGCGGC 54825
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 721 AATGAAACCACTACAAACCTGATTTCAAATGCGATGTACAGCATATTAGAAAACGCCAGGC 780
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54824 GACGAGACCAAGCGGATACCTTGACGGGTGGTACCAGCAGCTGTGCGCAACCGTGAC 54765
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 GTTTACGAGAACTGCGCAGCCATCCTGAACACTGATGCTCAGGAGTGGAGGAGCCTTG 840
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54764 CAGTGGGACCTGCTGCAGCGCGACCCGCTGCTGCTGCCGGGCGCATCGAGGAGATGCTA 54705
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 CGTTTCAGAGCGCGCGCCCGCTTTTGAGCGCATTTGCCAAGCGGATACGGAGATCGGG 900
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54704 CPTTGGACCGCCCGCTTAAAGAACATGTGCGCGGTGTGACCGCGGATACCGAGTTTCAC 54645
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 GGGCACTGATTAAGAAGTGATATGTTTTCGGCTTTCGGCATCGGCAATCTGTGAT 960
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54644 GGCACGGCTTGTGTCGCGCGAGAGATGATGCTCTCTTCAGATGCGGCACTTCGAC 54585
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 GAACAAAGTTTGACAGACCGCACATGTTTGTATATCGCGCCATCCCAATCCGCATATT 1020
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54584 GAGCGGCTTCTGTGAACCGGAAAGTTTGATGTTTCAGCGAAATCCAAACAGCCACTTG 54525
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1021 GCGTTTGGCACCGCATCTCAATTTTCCCTTGGGGCGCGCTTCCCGCTCTGA 1073
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54524 GCGTTTGGCTTCGGCACGCAATTTCTGCTGGCAATCAGCTGGCCCGGTTGA 54472
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AAI99683_39/c
Continuation (40 of 44) of AAI99683 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
```

```
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Query Match 9.8%; Score 118.2; DB 4; Length 110000;
Best Local Similarity 48.1%; Pred. No. 5.9e-26;
Matches 401; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

QY 241 ATCAATTAACATGACCCGCGAGCATACAAATAATCCGTCAGTCGTCGACAAAGCCCTTT 300
DB 48025 ATGATCGACATGACGATCCGACATCTGTCGGCGCAAGCTGTTAAACGCGGCTTC 47966

QY 301 ACTCCGCGGTGATGACATGGACCGAGATTCAGAAATCAAGAAATCAAGATGACTGATT 360
DB 47965 ACCCGAAGCGGTGAAGGACAAGGAGCGGTGATGCGCGCTGTGTGACACCCCTGATC 47906

QY 361 CAAATAATTCAGGGCGCCAGTGTGACCTGCTTCACGATTTTTCATACCCGCTTCGG 420
DB 47905 GAGCCCTGTGCGAACCGCGCGAGTGTGACTTCGTGGGACCTGGCGCGCTACCG 47846

QY 421 GTTATTGTGATATCTGAGCTGTCTGGGAGTGCCTTCAGCGCAGATGGAACAGTTTAAAGCA 480
DB 47845 ATGCGGTGATCGCGCATCTCGGGTGGCTCCAGACGACGCGGACATGTTCTTCGGG 47786

QY 481 TGCTGTGATCTCTGTGTGCTGATGACGAGGATAAAGTGAAGAGCTGAAAGACCTTT 540
DB 47785 TGGTCCGACGATCTGGTGA - CATTCCTAGTTCGATGTCTCAAGAGGATTTCCAGAT 47727

QY 541 TTGGAAGACGAGATAAGTGTGAGGAAGAACTGGCCGCGTTTTCGCGGATCATAGAA 600
DB 47726 CACCATGACGCTTCGCGGCTACACGACTTCACCGCG - - - - -GCCACCATTCGC 47675

QY 601 GAAAGCGAACAACCGGACAGATATATTTCTATTTTACTGGAAGCGGAGAAACA 660
DB 47674 GCACGGCGCGGACCCCAACCGACGACTGTGTGACGCTGGTGGTTCGGAAGTTGAC 47615

QY 661 GCGGAGAGCTGTCCGTTGAAGAGCTGATTCGCTTTTGCACGCTGCTGCTGTGGCGGGA 720
DB 47614 GCGAGCGGCTTAAGCGACGACGAGCTGTGTATGAGAGCGCTGTGATCTGATCGCGGC 47555

QY 721 AATGAAACCACTACAAACCTGATTTCAATGCGATGTACAGATATTAGAAACGCGAGGC 780
DB 47554 GACGAGACCAACGCGCATCTTCGAGCGGTGATACCGAGAGCTGTGTGCGCAACGCTGAC 47495

QY 781 GTTTACGAGGAACCTCGGACGACCTCCTGAACTGATGCTTCAGCGAGTGGAGACCTTGG 840
DB 47494 CAGTGGACCTCTGCTGACGCGGACCCCGCTGTGTGCTGCGCGGCGCATCGAGAGATGCTA 47435

QY 841 CGTTTCAGAGCGCGCGCCCGTTTGAAGCGCATTCGCAAGCGGATACGAGATCGG 900
DB 47434 CGTTGGACCGCCCGGTAAAGAAATGTGCCGGTGTGTGACCGGATACGAGTTTCAC 47375
```

```
QY 901 GGCACCTGATTAAGAGGTGATGATGTTTGGCGTTTGGCATCGGCAATCGTAT 960
DB 47374 GGCACGCGCTTGTGTGCGCGGAGAGATGATGCTGCTCTTCGAGTCGGGAACTTCGAC 47315

QY 961 GAAGCAAGGTTTCACAGACGCGCATGTTTGATATCCGCGCATCCCAATCCGCATATT 1020
DB 47314 GAGCGCGTTTCTGTGAAACCGGAAAGTTTGATGTTTCAGGAAATCCAAACAGCCACTTG 47255

QY 1021 GCGTTTGGCCACGCGCATCCATTTTTCCTTTGGGCGCCCGCTTGCCTCTTGA 1073
DB 47254 GCGTTTGGCTTCGCGACGATTTCTGCTCGCAATCAGTGTGCGCGCTTGA 47202
```

RESULT 13

AAI99682_08
Continuation (9 of 45) of AAI99682 from base 800001 (Mycobacterium tuberculosis strain H:
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 9.8%; Score 116.4; DB 4; Length 110000;

Best Local Similarity 55.4%; Pred. No. 2.3e-25;

Matches 225; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 659 GCTGTCCGGTGAAGAGCTGATTCGCTTTTGACGCTGCTGTGTGGCGGAAATGAAC 728

DB 72139 GCTGTCCGACCGCACTGTACTTCTTCCACTACTGTTTCAGCGCGCGCGCAAC 72198

QY 729 CACTACAAACCTGATTTCAAAATGCGATGTACAGCATATTAGAAACCCAGGCGTTTACGA 788
DB 72199 CACCCGTAACCTCAATTCGCGGGGCTGCTGGCGCGAGAACCTCTGACCACTGCA 72258
QY 789 GGAATCTGCGCAGCCATCTCAACTGATGCTCAGGCAAGTGGAGAGCTTGCGTTTCAG 848
DB 72259 AACGCTGCGAAGCGATTTTGAGTTGTTGCGCACTGCGATCGAAGAGATCTGAGGTGGAC 72318
QY 849 AGCGCGCGCCCGGTTTTCAGGCGCATTCGCAAGCGGATACGAGATCGGGGGGCACTT 908
DB 72319 GTCGCGCTCACCATCGAAGCGGCGACGGGTCCTGCGGTTCAGCTGGCGGCCAGCC 72378
QY 909 GATTAAAGAGTGTATATGTTTGGCGTTTGTGGCATCGCGCAATTCGCGATGAGCAACAA 968
DB 72379 GATCGAGCGGGTTCAGAGGTTGTGTTGGAGAGGCTCGCGCAACCTGATCCAGCCT 72438
QY 969 GTTTGACAGACCGCATCTTTGATATCGCGCCATCCCAATCCGCATATTCGGTTTG 1028
DB 72439 GTTCGACCGCGGAGAGTTGATATCACCCGAAACCAATCCGACCTGGGTTTCGG 72498
QY 1029 CCACGCGATCCATTTTTCGCTTGGGCGCCCGCTTGCCTGCGCTTTGAA 1074
DB 72499 TCAGGGGTGCACTATTGCTGGCGCCATCTGGCTCGCTCGAA 72544

RESULT 14

AAI199683_08
Continuation (9 of 44) of AAI199683 from base 800001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments LOCUS AAI199683 Accession Aai199683

Fragment Name	Begin	End
WP AAI199683_00	1	110000
WP AAI199683_01	100001	210000
WP AAI199683_02	200001	310000
WP AAI199683_03	300001	410000
WP AAI199683_04	400001	510000
WP AAI199683_05	500001	610000
WP AAI199683_06	600001	710000
WP AAI199683_07	700001	810000
WP AAI199683_08	800001	910000
WP AAI199683_09	900001	1010000
WP AAI199683_10	1000001	1110000
WP AAI199683_11	1100001	1210000
WP AAI199683_12	1200001	1310000
WP AAI199683_13	1300001	1410000
WP AAI199683_14	1400001	1510000
WP AAI199683_15	1500001	1610000
WP AAI199683_16	1600001	1710000
WP AAI199683_17	1700001	1810000
WP AAI199683_18	1800001	1910000
WP AAI199683_19	1900001	2010000
WP AAI199683_20	2000001	2110000
WP AAI199683_21	2100001	2210000
WP AAI199683_22	2200001	2310000
WP AAI199683_23	2300001	2410000
WP AAI199683_24	2400001	2510000
WP AAI199683_25	2500001	2610000
WP AAI199683_26	2600001	2710000
WP AAI199683_27	2700001	2810000
WP AAI199683_28	2800001	2910000
WP AAI199683_29	2900001	3010000
WP AAI199683_30	3000001	3110000
WP AAI199683_31	3100001	3210000
WP AAI199683_32	3200001	3310000
WP AAI199683_33	3300001	3410000
WP AAI199683_34	3400001	3510000
WP AAI199683_35	3500001	3610000
WP AAI199683_36	3600001	3710000
WP AAI199683_37	3700001	3810000
WP AAI199683_38	3800001	3910000
WP AAI199683_39	3900001	4010000
WP AAI199683_40	4000001	4110000
WP AAI199683_41	4100001	4210000
WP AAI199683_42	4200001	4310000

WP	AAI199683_43	4300001	4403765
Query Match 9.8%; Score 116.4; DB 4; Length 110000;			
Best Local Similarity 55.4%; Pred. No. 2.3e-25;			
Matches 225; Conservative 0; Mismatches 181; Indels 0; Gaps 0;			
QY	669	GCTGTCGGGTGAAGAGCTGATTCGGTTTTCAGCGTCTGCTGCTGGTGGCGGAAATGAAC	728
DB	74269	GCTGTCGGACGCGGAATCTGACCTGTTCTTCATCTACTGTTCTAGCGCGCGGGAAC	74328
QY	729	CACTACAAACCTGATTTCAAAATGCGATGTACAGCATATTAGAAACCGCGGCGTTTACGA	788
DB	74329	CACCCGTAACCTCAATTCGCGGGGCTGCTGGCGTGGCGAGAACCTTGACCACTGCA	74388
QY	789	GGAATCTGCGCAGCCATCTCAACTGATGCTCAGGCAAGTGGAGAGAGCTTGCGTTTCAG	848
DB	74389	AACGCTGCGAAGCGATTTTGAGTTGTTGCGCACTGCGATCGAAGAGATCTGAGGTGGAC	74448
QY	849	AGCGCGCGCCCGGTTTTCAGGCGCATTTGCCAAGCGGACACGAGATCGGGGGGCACTT	908
DB	74449	GTCGCGTTCACCATCGAAGCGGCGACGGGTCCTGCGGTGCGGTGAGCTGGCGGCGAGCC	74508
QY	909	GATTAAAGAGTGTATATGTTTGGCGTTTGTGGCATCGGCAATCCCAATCCGCATATTCGGTTGG	968
DB	74509	GATCGAGCGGGTCAGAAAGTTTGTGGTGGAGGGCTCGGCCAACCGTGATCCAGCGT	74568
QY	969	GTTCGACAGCCGACATGTTTGATATCCGCGCGCATCCCAATCCGCATATTCGGTTGG	1028
DB	74569	GTTTCGACCGCGGACAGTTCGATATCACCCGAAACCAATCCGACCTGGGTTTCGG	74628
QY	1029	CCACGCGATCCATTTTTCGCTTGGGCGCCCGCTTGCCTGCGCTTTGAA	1074
DB	74629	TCAGGGGTGCACTATTGCTGGCGCCATCTGGCTCGCTCGAA	74674
RESULT 15			
ACF06136			
ID	ACF06136 standard; DNA; 1248 BP.		
XX	AC ACF06136;		
DT	30-SEP-2003 (first entry)		
DE	Bacterial P450 enzyme encoding DNA SEQ ID NO:55.		
XX	Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;		
XX	diol; alkene; chirality; thermostability; gene; ds.		
OS	Eubacteria.		
XX	Key Location/Qualifiers		
FT	1..1248		
FT	/*tag= a		
FT	/product= "P450 enzyme"		
XX	W02003052050-A2.		
XX	26-JUN-2003.		
XX	05-AUG-2002; 2002WO-US024910.		
XX	03-AUG-2001; 2001US-0309497P.		
XX	(DIVE-) DIVERSA CORP.		
XX	Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short C;		
XX	WPI; 2003-541641/51.		
XX	P-PSDB; ABR82146.		
XX	Novel polypeptide having P450 enzyme activity and polynucleotides		
XX	encoding the polypeptide, useful for catalyzing the hydrolysis of		
XX	epoxides and arene oxides to their corresponding diols.		

XX

PS Claim 18; Page 364; 365pp; English.

XX

CC The present sequence encodes a bacterial p450 enzyme. p450 enzymes can be used to catalyze the hydrolysis of epoxides and arene oxides to their corresponding diols. p450 enzymes can also be used for hydrolysing an alkene, for producing a compound of a desired chirality, and for increasing thermotolerance or thermostability of a p450 polypeptide

XX

SQ Sequence 1248 BP; 127 A; 451 C; 510 G; 160 T; 0 U; 0 Other;

Query Match 9.6%; Score 114.2; DB 8; Length 1248;

Best Local Similarity 46.7%; Pred. No. 8.6e-26;

Matches 399; Conservative 0; Mismatches 453; Indels 3; Gaps 1;

QY 230 TTGGAATTCATCATTAACATGACCGCGGAGCATACAAAATCCGTTTCAGTCGTGA 289

DB 263 TCAGAACTGGCTGCTTCTCTGACCCGCGCCCGACACGGGCTGCGGCCAGGTGG 322

QY 290 ACAAGCCTTTACTCCGCGGTGATGAAGCAATGGGAACCGAGAATTCAGAATACAG 349

DB 323 CCGCGCGGTGAGCGCTCGGGCGGTACGGCCCTTGGCTCCCGGTACCGGAGATCGCG 382

QY 350 ATGAACGTATCAAAATTCAGGGGCGAGTGAGTTGACCTTGTTCAAGATTTTCAT 409

DB 383 AGGAATCTGTACGGCCCTCGCGCGCGCCCGCTGGTCTGAATGGTCGAGGGGCTTCGCG 442

QY 410 ACCCGCTTCGGGTATTGTGATATCTGAGCTGTGGGAGTGCCCTTCAGCGCAGATGAAC 469

DB 443 CCGCGTTCCCGCTGTGCTGCTGCGGGGCTGCTCGGGGTGACCCCGGTGGTGCCGT 502

QY 470 AGTTAAAGCATGCTCTGATCTTCTGCTAGTACACCGAAGGATAAAGTGAAGAAGCTG 529

DB 503 GGTTCGGGAGGAGGCG---CTGCCCTCCAGCGGTCCGGCGGACCCAGGGGACCGCGT 559

QY 530 AAAAGCCTTTTGGAGAACGAGATAAGTGTGAGGAGAACTGGCGCGCTTTTTCGCG 589

DB 560 CGCGCGCGCGCTCGCACGGGCGCGACCGCGCGCGCGACCTTGACGCGTACTTCGCGG 619

QY 590 GCATCATAGAGAAAGCGAAACAAACCGGAACAGAGATATTATTTCTATTATTAGTGAAG 649

DB 620 CGAGCTGGCGCGCGCGCTCGAGGACCGGGGGACCTGCTCTCGCGCTGCGCGGG 679

QY 650 CGGAAGAAACAGCGGAGAACTGTCCGTTGAAGAGCTGATTCGTTTTTGACGCTGCTGC 709

DB 680 CGCGCGCGGAGACCCCTCGCTGGGAGACGACCGCGCTGACCTGACCTGCTCCACCTCC 739

QY 710 TGGTGGCGGAATGAACCACTACAAACCTGATTTCAAATGGCATGTACAGCATATTAG 769

DB 740 TGACGGCGGGGACGAGACGACCGGCGCTGCTGGGCGAGGCGGTGCTGCGCTGCTGG 799

QY 770 AAACGCCAGCGGTTTACAGGAACTGGCGAGCCATCTTGAACTGATGCTCAGGCACTGG 829

DB 800 CGCGCGCGAGGTGGCGGAGGAACTGCGCGGACCCCGGCGCTGTTGCCGAACCGCGTGG 859

QY 830 AGGAGCCTTGGCTTTCAGAGCGCGCGCGCGCTTTTGGAGCGCATTTGCCAAGCGGATA 889

DB 860 ACCAGTTCTCTCGCGCACGACCCACCGTGCAGATGGTCAACCGGTGGCGCGCGGACG 919

QY 890 CGGAGATCGGGGGCACCTGATTTAAGAGGTGATGATGTTTGGCGGTTTGTGGCATCGG 949

DB 920 CGGAGCTGGCGCGCGGCGAGCGCTCCGCGGGGCGACCGGCTCCAGCTGCTGCTGGGTCGG 979

QY 950 CAATCTGTGATGAAGCAAGTTTGACAGACCGCACATGTTTGATATCCGCGCCATCCCA 1009

DB 980 CCCACCGGACCCCGCGCTTCCGCGACCCCGACCGGCTGGACATCCGCGGGACACCG 1039

QY 1010 ATCCGCATATTGGTTGGCCACCGGCAATCCATTTTGTGCTTGGGGCCCCGCTTGGCCGTC 1069

DB 1040 GCCGCGCATGCGCGTTCGGTCTGCGCATCCACTACTGCTGGGGCGCGGCTGGCACGCG 1099

QY 1070 TTGAAGCAATATCG 1084

DB 1100 CGGAGCGGAGATCG 1114

Search completed: May 28, 2004, 20:33:07

Job time : 406 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 28, 2004, 20:08:15 ; Search time 2408 Seconds
(without alignments)
14769.874 Million cell updates/sec

Title: US-09-869-334B-2

Perfect score: 1191

Sequence: 1 atgaatgtgttaaacgcgcg.....gcttcggtgtgaaatgtaa 1191

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106.2	8.9	975	28	BZ675691 PUBDH78TD
2	81.2	6.8	1285	28	BZ567880 pacs2-164
3	77.4	6.5	479	14	CD091452
4	72.6	6.1	1127	28	BZ548452

5	53.6	4.5	330	28	AV174775
6	53.4	4.5	636	14	CD486042
7	52	4.4	1223	28	BZ551987
8	49.4	4.1	638	14	CD486072
9	44.2	3.7	1101	28	CNS0039G
10	43.4	3.6	588	28	AZ934336
11	42.8	3.6	1136	28	BZ563995
12	42.8	3.6	1201	13	BX381961
13	42.2	3.5	718	29	CG092084
14	40.6	3.4	432	13	BX577443
15	40.6	3.4	1201	13	BX406481
16	40.6	3.4	1201	13	BX417669
17	40.6	3.4	1201	13	BX442405
18	40	3.4	244	14	CF050066
19	39.8	3.3	2096	11	AY109482
20	39.6	3.3	591	14	CB923439
21	39.6	3.3	601	28	BZ374029
22	39.4	3.3	827	29	CG041562
23	39.4	3.3	1201	13	BX361080
24	39.4	3.3	1201	13	BX376097
25	39.2	3.3	575	12	BG515155
26	39.2	3.3	654	9	AL888423
27	39.2	3.3	658	9	AL862443
28	39.2	3.3	671	9	AL886494
29	39	3.3	1091	13	BX424950
30	38.8	3.3	469	9	AA246005
31	38.8	3.3	636	14	CD429048
32	38.8	3.3	641	12	BM329835
33	38.6	3.2	952	29	CNS0605L
34	38.6	3.2	1023	28	CC292361
35	38.6	3.2	1057	28	CC293665
36	38.4	3.2	247	14	CK228885
37	38.4	3.2	268	14	CK228977
38	38.4	3.2	306	14	CK218284
39	38.4	3.2	503	14	CK228926
40	38.4	3.2	523	14	CK228927
41	38.4	3.2	552	14	CK228929
42	38.4	3.2	563	14	CK228925
43	38.4	3.2	593	9	AW111487
44	38.4	3.2	595	14	CK228997
45	38.4	3.2	624	14	CK228974

ALIGNMENTS

RESULT 1
BZ675691
LOCUS
DEFINITION PUBDH78TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa029M12,
Genomic survey sequence.
ACCESSION BZ675691
VERSION BZ675691.1 GI:28226790
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

BZ675691 975 bp DNA linear GSS 05-FEB-2003
PUBDH78TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa029M12,

REFERENCE
AUTHORS
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
1 (bases 1 to 975)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw
TIGR Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: Sheared ends.

REFERENCE
AUTHORS
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
1 (bases 1 to 975)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw
TIGR Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: Sheared ends.

REFERENCE
AUTHORS
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
1 (bases 1 to 975)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw
TIGR Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: Sheared ends.

REFERENCE
AUTHORS
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
1 (bases 1 to 975)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw
TIGR Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: Sheared ends.


```

University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. .1285
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_7256"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library"

```

ORIGIN	Query Match	6.8%; Score 81.2; DB 28; Length 1285;
	Best Local Similarity 50.7%; Pred. No. 1.6e-12;	
	Matches 222; Conservative 0; Mismatches 213; Indels 3; Gaps 1;	
QY	641 TAGTGGACGGAGAAACACAGCGGAGAGACTGTCGGTGAAGAGCTGATTCGGTTTGGCA 700	
Db		
QY	111 TGGTGCAGGCCCGCACCGACGAGAGCGGCCAGTTGAGCGAAGCGGAACTCGTCTCCATGGCCC 170	
Db		
QY	701 CCGTCTGCTGCTGGCCGGAATGAACCACTACAAACCTGATTTCAATGCGATGTACA 760	
Db		
QY	171 ACCTGCTGATGATGACGGCTTCGAGACCAACCATGAACATGATGCGCAACGGCGTGCTCA 230	
Db		
QY	761 GCATATTAGAAACGCCGAGCGCTTTACGAGAGAACTGCGACGCCATCTCGAACTGATGCGCTC 820	
Db		
QY	231 CCGTCTGCTGAACCCGGAGCAACTGCGCTTCGTCGGGGCGCAGCGGAACTCCTGCCCA 290	
Db		
QY	821 AGCGAGTGGAGGAA--GCCTTCGGTTTCAGAGCGCGCGCCCGGTTTTCAGCGCGGATTG 877	
Db		
QY	291 ACGCCATGGAAGAACTGCTGCCACGACAGTCCGGTGGCGGCTCGATGTTTCGCGTTTCA 350	
Db		
QY	878 CCAAGCGGATACGAGATCGGGGGGCACCTGATTAAGAAAGTGATATGGTTTTCGGCT 937	
Db		
QY	351 CCGTGGAAACGTGGAACATGACGCGGTCACCATTTCCCGCCGCGAATAATCATCTGGTCT 410	
Db		
QY	938 TTGTGSCATCGGCAATCGTGATGAAGCAAGTTTGACAGACCGCACATGTTTGATATCC 997	
Db		
QY	411 CGAACCTGACCGCCACACACGATCGCGGTTGAGCATCCGACCGCTTCGACCTCA 470	
Db		
QY	998 GCGGCCATCCCAATCGCATATTTGGCTTTTGGCCAGCGCATCCATTTTTCCTTTGGGGCCC 1057	
Db		
QY	471 CCGCGAACACCGATGCGCATCTCGGCTACGGCTTCGGCGTGCACTACTGCGTCGGCGCT 530	
Db		
QY	1058 CGCTGCGCGCTTCGAAG 1075	
Db		
QY	531 CGCTGCGCGCTGGAAG 548	
Db		

RESULT 3	CD091452	479 bp	mRNA	linear	EST 14-SEP-2003
LOCUS	CD091452/c				
DEFINITION	MC1-0087P-A156-F03-U.G MC1-0087 Schistosoma mansoni. cDNA clone				
ACCESSION	MC1-0087P-A156-F03.G, mRNA sequence.				
VERSION	CD091452				
KEYWORDS	CD091452.1 GI:34642252				
SOURCE	EST.				
ORGANISM	Schistosoma mansoni				
	Schistosoma mansoni				
	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;				
	Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.				
REFERENCE	1 (bases 1 to 479)				
AUTHORS	Verjovskij-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,				
	Ojopi,E.P.B., Paqueta,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,				
	Kitiiaima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,				
	Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,				
	Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,				

FEATURES	Location/Qualifiers
source	1..975
	/organism="Zea mays"
	/mol_type="genomic DNA"
	/strain="B73"
	/db_xref="taxon:4577"
	/clone="ZMMBTa029M12"
	/length="24 0.6 1.0 KB"
	/notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
	Cot selected genomic DNA library"
ORIGIN	
Query Match	8.9%; Score 106.2; DB 28; Length 975;
Best Local Similarity	51.8%; Pred. No. 4.2e-20;
Matches	268; Conservative 0; Mismatches 243; Indels 6; Gaps 1;
QY	564 GGAGAACTGGCGCGTTTTTCGGGCATCATAGAGAAAGCGAAGCAAAACCGGNACA 623
DB	157 GGCGAGCTGGCGGCATACATCAAAACCTGGTGGCGGAACGTGACGCGAATCACGCGA 216
QY	624 GGATATTATTCTATTATTAGTGAAGCGGAAGAAACAGCGGAGAGCTGTCGGTGAAGA 683
DB	217 TGACTTCATGACCGCCTGACCCAGCCGAGAAAGGCGAGTCCCTCAACGAGGAGCA 276
QY	684 GCTGATTTCGGTTTGGACGGTCTGCTGTGGCGGGAATGAAACCACTACAAACCTGAT 743
DB	277 ATTGCTGCTGTTGTGCAATTGCGATTCCGATCTCGTTCGCGGTTACGAAGGCAGTGCCTCGCAGAT 336
QY	744 TTCAAATGCGATGTACAGCATATTAGAAACGCCAGCGGTTTACAGGAGACTGCGCAGCCA 803
DB	337 CCCCACATTATCCAGGTGCTGTGGAACAACCCCGCGCAGTGGCAACAGCTCAAGGCCGA 396
QY	804 TCTGAACTGATCGCTCAGGCGCTGGAGGAAGCTTGCCTTTTCAGAGCGCGCGCCCGCG - 862
DB	397 TCCGAGCAGATACCGAGCGCTGCAAGAGCTCTGGCGTACATCCCGTGTGCCTCGGC 456
QY	863 -----TTTTCAGCGCGATTGCCAAGCGGGATACGAGATCGGGGGGACCTGATTAAAGA 917
DB	457 GGCGATGTTTGTGCACTACGCGCTGGAAGACATTCAGTTCGGTGAACCTCGTGGGCCA 516
QY	918 AGTGTATATGGTTTGTGGCGTTTGTGGCATTCGGCAATTCGTGATGAAGCAAGTTTGACAG 977
DB	517 GGCGCAGCGGGTGTTCGCTTCGATCGCGCGCGCAACCATGACCCGCGCCGCTTTGAAA 576
QY	978 ACOGCACATGTTTGATATCGCGCGCATCCCAATCCGCAATTGCGTTTGGCCACAGCAT 1037
DB	577 CCACAGGGCTGGACCTGCAACGGGATGCACGCGCCATTTCGGCTTTGGCCACAGGCT 636
QY	1038 CCATTTTTCGCTTTGGGCGCCCGCTTGCCTGCTTTGAA 1074
DB	637 GCATCACTGCATCGGCTCAGCCCTCGCGCGGTCGAA 673

RESULT 2	BZ567880	LOCUS	pacS2-164_7256.x1	1285 bp	DNA linear	GSS 17-DEC-2002
DEFINITION	pacS2-164_7256, genomic survey sequence.					
ACCESSION	BZ567880	VERSION	BZ567880.1	GI:27199936		
KEYWORDS	GSS.					
SOURCE	Pseudomonas aeruginosa					
ORGANISM	Pseudomonas aeruginosa					
REFERENCE	1 (bases 1 to 1285)					
AUTHORS	Spencer, D.H., Ray, R.D., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.					
TITLE	Whole-Genome-Sequence variation among multiple isolates of <i>Pseudomonas aeruginosa</i> library					
JOURNAL	J. Bacteriol. (2002) In press					
COMMENT	Contact: Chris K. Raymond Genome Center					

Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
 Sa.R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
 Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
 Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
 Transcriptome analysis of the acelomate human parasite Schistosoma
 mansoni
 Nat. Genet. 35 (2), 148-157 (2003)

22879926
 Contact: Dr. Sergio Verjovski-Almeida
 Departamento de Bioquímica
 Instituto de Química - Universidade de São Paulo
 Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
 Brasil

Tel: +55-11-3091-2173
 Fax: +55-11-3091-2186
 E-mail: verjovski@usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
 Genome Project. All sequences in the project were assembled and
 annotated. This entry and all the assembled sequences can be seen
 in the following URL <http://bioinfo.ig.usp.br/schisto/>
 Plate: MCI-0087P-A156 row: 3 column: F.

FEATURES

1..479
 Location/Qualifiers
 /organism="Schistosoma mansoni"
 /mol_type="mRNA"
 /db_xref="taxon:6183"
 /clone="MCI-0087P-A156-F03.G"
 /sex="mixed pool"
 /dev_stage="Cercaria"
 /lab_host="Biomphalaria glabrata"
 /clone_lib="MCI-0087"
 /note="Vector: pGEM T-easy"

ORIGIN

Query Match 6.5%; Score 77.4; DB 14; Length 479;
 Best Local Similarity 49.9%; Pred. No. 1.4e-11;
 Matches 223; Conservative 0; Mismatches 221; Indels 3; Gaps 1;
 QY 334 ATTCAGAAATCACAGATGAATCGATGTAATTAATAATTTACGGGGCGAGTGAGTTGACCTT 393
 |||||
 Db 456 ATTCAGAGATTGCTGAGCGCTTGATTGATGATGAGTGGCTAAGGGCGGATGGAATTT 397
 |||||
 QY 394 GTTCAGATTTCATACCGCTTCGGTTGTTGATATCTCAGCTGCTGGAGTGCCT 453
 |||||
 Db 396 ATTCGGAAATCGCCCTTCCTGTTCCGATTACAGTAAATTTGTAATGCTGGGATACCT 337
 |||||
 QY 454 TCAGGCGAGATCGAAACAGTTTAAAGCATGCTGATCTTCGTGCTAGTACACCGAAGGAT 513
 |||||
 Db 336 TCGGCTGACCAGAGAAATTCAGACAAATGCTCAACGCAATGACTCAGACCTTAGAGCCT 277
 |||||
 QY 514 AAAAGTGAAGAGCTGAAGAAAGCTTTTGGAGAAACGAGATAAGTGTGAGGAAGAACTG 573
 |||||
 Db 276 GSCAGCATCACAAATCCTTCAGTCTTAC---ACGTATCAAAATCAGCTGCTGATGAAGT 220
 |||||
 QY 574 GCCGGTTTTTTCGGGCATCATAGAAAGAAAGCGAAACAAACCGCAACGATATATT 633
 |||||
 Db 219 CATGACTATCTCGAACCGCTAGTGCAGGAACGCGAAAGAAATCCAAAGCGCGATCTAAT 160
 |||||
 QY 634 TCTATTTAGTGAAGCGGAAGAAACAGCGAGAGCTGTCCGGTGAAGAGCTGATTCGG 693
 |||||
 Db 159 ACAGCTCTGTTGAGCAGAAATGAGACAAATACTTGAAGATCAGAGTAAATGGA 100
 |||||
 QY 694 TTTTGACGCTGCTGCTGTTGGCCGGAATGAACCACTACAAACCTGATTTCAAAATGG 753
 |||||
 Db 99 AATATAATGCTTCTGCTAGTGCAGGGCAGACAAACCGTTAATTTAATTTGGCAACGGT 40
 |||||
 QY 754 ATGTACAGATATTAGAAAGCCGAGC 780
 |||||
 Db 39 ATGTTTGTCTTATGCAACATCCGGAC 13
 |||||

RESULT 4
 BZ548452

LOCUS
 DEFINITION BZ548452 1127 bp DNA linear GSS 17-DEC-2002
 pacsl-60.1124.s1 pacsl-60 Pseudomonas aeruginosa genomic clone
 pacsl-60.1124, genomic survey sequence.

ACCESSION BZ548452

VERSION BZ548452.1 GI:27152033

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1127)

AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
 Burns,J.L., Kaul,R. and Olsen,M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1..1127
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone="pacsl-60.1124"
 /clone_lib="pacsl-60"
 /note="clinical isolate 1-60 Whole genomic shotgun
 library."

ORIGIN

Query Match 6.1%; Score 72.6; DB 28; Length 1127;
 Best Local Similarity 57.1%; Pred. No. 5.7e-10;
 Matches 132; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
 QY 240 CATCATTAACATGACCGCGGAGCATACAAAATCCGTTTCAGTCTGTGAACAAAGCCTT 299
 |||||
 Db 250 CATGCTCACTCGACCGCGGACCATACCCGCTCGCTGCTGGCGCGGCTT 309
 |||||
 QY 300 TACTCGCGCTGATGAAGCAATGGGAACCGAGAAATTCAGAAATACAGATGACTGAT 359
 |||||
 Db 310 CACCCCGCGCTGAGCGGCTGCAACCGCATATACACGGATCACCGAGGAATTGCT 369
 |||||
 QY 360 TCAAAAATTCAGGGCGCAGTGAGTTGACCTTTTCAGATTTTTCATACCGGCTTC 419
 |||||
 Db 370 GGACGCGATGCGCGCGCGAACAGCGCGACCTGATGGCGGACTTCGCGATCCCGCTGAC 429
 |||||
 QY 420 GGTATTGTGATATCTCAGCTGCTGGAGTGCCTTCAGCGCAGATGGAACA 470
 |||||
 Db 430 CATCGCGGTGATCTTCAGCTGCTGGGCTATCCCGAGCGCGGCGGAACA 480
 |||||

RESULT 5

AY174775

LOCUS

AY174775 330 bp DNA linear GSS 13-JAN-2003
 AY174775 Mycobacterium avium subsp. paratuberculosis DNA
 Mycobacterium avium subsp. paratuberculosis genomic clone
 Mptb61.15, genomic survey sequence.

ACCESSION AY174775

VERSION AY174775.1 GI:27689314

KEYWORDS GSS.

SOURCE Mycobacterium avium subsp. paratuberculosis

ORGANISM Mycobacterium avium subsp. paratuberculosis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium;
 avium complex (MAC).

REFERENCE 1 (bases 1 to 330)

AUTHORS Nielsen,K.K. and Ahrens,P.

TITLE Subtractive hybridization and bioinformatics: combining genetic methods to obtain new information on species-specific elements of Mycobacterium avium subspecies paratuberculosis

JOURNAL COMMENT

Unpublished (2003)
Contact: Nielsen KK
Bacteriology
Danish Veterinary Institute
Bulowsvej 27, Copenhagen V, 1790, Denmark
Email: kni@veti.dk
Class: unknown.

FEATURES Location/Qualifiers
source
1..330
/organism="Mycobacterium avium subsp. paratuberculosis"
/mol_type="genomic DNA"
/sub_species="paratuberculosis"
/db_xref="taxon:1770"
/clone="pMtb61.15"
/clone_lib="Mycobacterium avium subsp. paratuberculosis DNA"

ORIGIN

Query Match 4.5%; Score 53.6; DB 28; Length 330;
Best Local Similarity 48.7%; Pred. No. 0.00017;
Matches 146; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

604 AAGCGAAACAAACCGAACAGATATTTCTATTTAGTGAAGCGGAAACACAGC 663
|||||
26 AAGCGCGGAGCCAGGTGACGATACGCACTTCTTACTCCAGCGCGAGGTGATGGT 85
|||||
664 GAGAGCTGTCGGTGAGAGCTGATTCGGTTTCACGCTGCTCTGGTGCCCGAAT 723
|||||
86 CAGAGCCTCAGCGATTTGGAATTTAACTGTCTTATGTTGCTGATCGATCGCGAGT 145
|||||
724 GAAACCACTACAAACCTGATTTCAAATCGGATGTACAGCATATTAGAAACGCGAGCGGT 783
|||||
146 GATAGACCCCAATCTGTAGACGGGAATCTTGCTCTGCTGAGCATCCCCAAGA 205
|||||
784 TACAGGAAGTGGCGAGCCATCTGAACTGATGCTCAGGAGTGGAGAGCCCTGGGT 843
|||||
206 CTGACGCGGCTTAAGGACAGACCCCTTCACTATCGCGACCGGATCGAAGAAATCTGGCG 265
|||||
844 TTCAGAGCGCGCGCCCGTTTGGAGCGCATTCGCAAGCGGATACGAGATCGGGGG 903
|||||
266 TACACGTCGCGGTGACGGCATCTCTCGCACCGCGCAAGACACTGAGTTGCGCGGG 325
|||||

RESULT 6

CD486042 636 bp mRNA linear EST 01-JUL-2003
LOCUS
DEFINITION
CFUS5.3A03 Cotton Root and Hypocotyl tissues infected with Fusarium oxysporum f.sp. vasinfectum (Fov) Lambda ZIPLOX Library (CFUS)
Gossypium hirsutum cDNA clone CFUS5.3A03 5' similar to Chain A,
X-Ray Structure Of Nitric Oxide Reductase, mRNA sequence.

CD486042
CD486042.1 GI:31407007

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Gossypium hirsutum (upland cotton)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Caltiriona Dowd, Helen McPadden
Commonwealth Scientific and Industrial Research Organisation
Division of Plant Industry
Black Mountain Laboratories, Cnr Clunies Ross Street & Barry Drive,
Black Mountain, Canberra, ACT, 2601, Australia
Tel: 61 2 6246 4914, 6246 5377
Fax: 61 2 6246 5000

Email: Caltiriona.Dowd@csiro.au, Helen.McPadden@csiro.au
Vector clipped sequences Bases 1-17 (GTGAGCCACGCGTCCG): Sali
adapter

Seq primer: M13 reverse primer

High quality sequence stop: 636.

FEATURES

source

1..636
/organism="Gossypium hirsutum"
/mol_type="mRNA"
/cultivar="DeltaEMERALD"
/db_xref="taxon:3635"
/clone="CFUS5.3A03"
/tissue_type="Fov infected root and hypocotyls tissues"
/dev_stage="seedlings, pooled from 5 hours to 7 days post infection with Fov."
/lab_host="V1090(ZL)"
/clone_lib="Cotton Root and Hypocotyl tissues infected with Fusarium oxysporum f.sp. vasinfectum (Fov) Lambda ZIPLOX Library (CFUS)"
/note="Vector: Lambda ZIPLOX; Site 1: Sali; Site 2: NotI; mRNA was prepared from pooled root and hypocotyl tissues of the cotton cultivar DeltaEMERALD at different stages of Fov infection from 5 hours to 7 days post infection. cDNA was synthesised from a NotI-oligo dT primer/adaptor using the manufacturers protocols (Life Technologies) and then ligated to a Sali adapter to facilitate directional cloning. The cDNA was cloned into the Sali and NotI sites of the Lambda ZIPLOX phage vector (Life Technologies). Constructed by Caltiriona Dowd and Helen McPadden."

ORIGIN

Query Match 4.5%; Score 53.4; DB 14; Length 636;
Best Local Similarity 49.0%; Pred. No. 0.00027;
Matches 172; Conservative 0; Mismatches 176; Indels 3; Gaps 1;

690 TCGGTTTGACGCTGCTGCTGGTGGCGGGAATGAACCACTACAAACCTGATTTCAA 749
|||||
13 TCCAGATTGCTTGTGCTTCTGCTGGCGGCAACCTCCTATGTCGCAATCATGAGCCCT 72
|||||
750 TCGCATGTACAGCATATTAGAAACCGCAGGGGTTTTCAGAGGAATCGCGAGCATCTG 809
|||||
73 GGGCTGCAACTCTGGCTCAGCACCTGATCAGCTGGCGCACTCAAGGCCAACCCATC 132
|||||
810 ACTGATGCTCAGGAGTGGAGAGAGCCCTTGGCTT---TCAGAGCGCGCGCCCGGTTTT 866
|||||
133 GCTTGGCGCCAGTTTGTGTAGAGAGCTCTGTGCTTACCACACTGGCTCTGCACTGCCAT 192
|||||
867 GAGGCGCATTCGCAAGCGGATACGAGATCGGGGGCACCTGATTAAAGAAAGGTGATAT 926
|||||
193 CAAGGCACTGCCAAGGAGGATGTGATGTCGCGCAAGCTTCGTCGGTGCAGAGG 252
|||||
927 GGTTTTGGCGGTTGTGGCATCGCAAAATCGTGATGAAGCAAGAGTTTTCAGACCCGCAT 986
|||||
253 CATTATGCAATCCCAACCACTCAGGCAACAGGAGCAAGAAAGTCTTCGAGAAATCTGTAT 312
|||||
987 GTTTCATATCGCGCGCATTCCTCAATCCCATATTGCGTTTGGCCACGGCAT 1037
|||||
313 GTTCAACATGAACCGCAAGTGGCCTTACTCAAGATCCTCTTGGCTTTGGCTT 363
|||||

RESULT 7

BZ551987/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 1223)

REFERENCE

1 (bases 1 to 1223)

REFERENCE

1 (bases 1 to 1223)

REFERENCE

1 (bases 1 to 1223)

BZ551987 1223 bp DNA linear GSS 17-DEC-2002
pacsl-60.3425.xl pacsl-60 Pseudomonas aeruginosa genomic clone
pacsl-60.3425, genomic survey sequence.

BZ551987.1 GI:27155901

GSS.

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1223)

AUTHORS

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE

Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library

JOURNAL

J. Bacteriol. (2002) In press

COMMENT

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

FEATURES

source

Location/Qualifiers

1. .1223

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"

/strain="1-60"

/db_xref="taxon:287"

/clones="pacs1-60_3425"

/clone_lib="pacs1-60"

/note="clinical isolate 1-60 Whole genomic shotgun
library."

ORIGIN

```

Query Match      4.4%; Score 52; DB 28; Length 1223;
Best Local Similarity 50.8%; Pred. No. 0.00093;
Matches 124; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 209 TSCGGAGGACGACGAGCTTATTTGGAATTCATCAATGAATGACCCGCGGAGCATA 268
Eb 534 TGAACCGCTGGCGGCTTCTACGACAACTTCATGCTGTTCAACGACGCGCGCGGCACA 475
QY 269 CAAAAATCCGTTGAGTCTGTAACAAAGCCTTTACTCCGCGGCTGATGAAGCAATGGGAAC 328
Db 474 CCGCGCTGGCAGGCTGTTGCTCGCGCTTCGCGCGCGCGCGCTGGGAG 415
QY 329 CGAAGATCAAGAAATCAGATGAACTGATTTCAAAATTTACGGGCGCAGTGTG 388
Db 414 CGCGCATCGAGGTCTGCTGGAAGAACTCTCTCGACAGCTTCTGAGCGCGCGGAGCCG 355
QY 389 ACCTGTTTCAGATTTTATACCCGCTTCGGTATTTGATATCTGAGCTGCGGAG 448
Db 354 ATCTGCTCAGGATTTCCGCGAACCGCTGAGATTCGCGGTGGCGCGAGCTGTCGCT 295
QY 449 TGCC 452
Db 294 TTCC 291

```

RESULT 8

CD486072

LOCUS

CD486072 638 bp mRNA linear EST 01-JUL-2003
CFUS5.4D09 Cotton Root and Hypocotyl tissues infected with Fusarium
oxysporum f.sp. vasinfectum (Fov) Lambda ZIPLOX Library (CFUS)
Gossypium hirsutum cDNA clone CFUS5.4D09 5' similar to Chain A,
X-Ray Structure Of Nitric Oxide Reductase, mRNA sequence.

ACCESSION

CD486072

VERSION 1 GI:31407037

KEYWORDS

EST.

SOURCE

Gossypium hirsutum (upland cotton)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.

1 (bases 1 to 638)

Dowd, C., Wilson, I., and McFadden, H.

Different Gene Expression Responses in Cotton Root and Hypocotyl

tissues during infection with Fusarium Wilt Disease

Unpublished (2003)

JOURNAL

Contact: Cairiona Dowd, Helen McFadden

Commonwealth Scientific and Industrial Research Organisation

Division of Plant Industry

Black Mountain Laboratories, Cnr Clunies Ross Street & Barry Drive,
Black Mountain, Canberra, ACT, 2601, Australia
Tel: 61 2 6246 4914, 6246 5377
Fax: 61 2 6246 5000

Email: Cairiona.Dowd@csiro.au, Helen.McFadden@csiro.au
vector clipped sequences Bases 1-17 (GTGACCCACGGCTCG): Sall
adapter

Seq primer: M13 reverse primer

High quality sequence stop: 638.

Location/Qualifiers

1. 638

/organism="Gossypium hirsutum"

/mol_type="mRNA"

/cultivar="DeltaEMERALD"

/db_xref="taxon:3635"

/clone="CFUS5.4D09"

/tissue_type="Fov infected root and hypocotyls tissues"

/dev_stage="seedlings, pooled from 5 hours to 7 days post

infection with Fov."

/lab_host="Y1090(ZL)"

/clone_lib="Cotton Root and Hypocotyl tissues infected

with Fusarium oxysporum f.sp. vasinfectum (Fov) Lambda

ZIPLOX Library (CFUS)"

/note="Vector: Lambda ZIPLOX; Site 1: Sall; Site 2: NotI;

mRNA was prepared from pooled root and hypocotyl tissues

of the cotton cultivar DeltaEMERALD at different stages of

Fov infection from 5 hours to 7 days post infection. cDNA

was synthesised from a NotI-oligo primer/adaptor using

the manufacturers protocols (Life Technologies) and then

ligated to a Sall adapter to facilitate directional

cloning. The cDNA was cloned into the Sall and NotI sites

of the Lambda ZIPLOX phage vector (Life Technologies).

Constructed by Cairiona Dowd and Helen McFadden."

ORIGIN

```

Query Match      4.1%; Score 49.4; DB 14; Length 638;
Best Local Similarity 46.4%; Pred. No. 0.0042;
Matches 198; Conservative 0; Mismatches 226; Indels 3; Gaps 1;

QY 614 AACCGAACAGGATATTTCTATTTTAGTGAAGCGGAAGAACACGCGGAGAGCTGT 673
Db 91 AACCGAGGATGATATCATAGTAAGCTCTGCACCTGACCAAGTCAAGCTGGAATATCG 150
QY 674 CCGGTGAAGAGCTGATTCGGTTTTCACGCTCTCTGCTGGCGGAAATGAACCACTA 733
Db 151 ACAAGTCGAGCGCTGTCCAGATTGCTTCTGCTCTCTGCTGGCGGCAACGCTACATGG 210
QY 734 CAACCTGATTTCAATGCGGATGTACAGCATATTAGAAACGCCAGCGCTTTACGAGAAC 793
Db 211 TCAACATGATAGCCCTGGGGGTGCGAACTCTGGCTCAGCACCCCTGATCAGCTGGCGAAC 270
QY 794 TCGCAGCCATCTCGAACTGATGCTCAGGAGTGGAGAGAGCTTGCGTT---TCAGAG 850
Db 271 TCAGGCCAACCCATCGTTGGGCCCGAGTTTGTGTGAGAGCTCTGTGCTTACCACATCG 330
QY 851 CGCGGCGCCCGGTTTTCAGGCGCAATTCACAGCGGATACGAGATCGGGGGGACCTGA 910
Db 331 CGTCTGCACTCGCCATCAAGCGCACTGCGCAAGGAGGATGTCATGATCGGCGACAGCTCG 390
QY 911 TTAAGAAGGTGATATGGTTTGGCGTTTGTGGCATGCGGCAATCGTGATGAAGAAAGT 970
Db 391 TCCGTGCAACGAGAGGATATTCGCATCCAACCCAGTCAGGCAACGAGGACGAAGAGTCT 450
QY 971 TTGACAGACGCGACATGTTTGATATCGCGCCCATCCCAATCCGCATATTCGTTTGGCC 1030
Db 451 TCGAGATCTCTGATGAGTTCAACATGAACCGCAAGTGGCCTTACTCAAGATCCTCTTGCT 510
QY 1031 ACGGCAT 1037
Db 511 TTGGCTT 517

```

RESULT 9

```

CNS00396/c
LOCUS       CNS00396             1101 bp      DNA          linear      GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
              BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
VERSION     AL063921
KEYWORDS    AL063921.1  GI:4941778
SOURCE      GSS.
ORGANISM    Drosophila melanogaster (fruit fly)
            Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
            Direct Submission
            AUTHORS   Not Available
            TITLE     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            JOURNAL   BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
            COMMENT   Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osogawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     1..1101
        /organism="Drosophila melanogaster"
        /mol_type="genomic DNA"
        /db_xref="taxon:7227"
        /clone="BACR08K10"
        /clone_lib="RPCI-98"
        /note="end : TET3"

ORIGIN
Query Match      3.7%; Score 44.2; DB 29; Length 1101;
Best Local Similarity 18.2%; Pred. No. 0.2;
Matches 45; Conservative 117; Mismatches 85; Indels 0; Gaps 0;

QY      512  ATAAAGTGAAGAGCTGAAAGACCTTTTGAAGACGACGATAGTGTGAGGAGAAC 571
DB      848  ADDWADWADWAAWAKWDDAWAWGARTARDRDWGDWDRAGKRGKRRDRKRADKDDAAD 789
QY      572  TGGCCGCGTTTTTTCGCGCATCATAGAGAAAGCGAAACACGACACGAGTATTA 631
DB      788  DRDDAATWTTTTRDTRDDWKKWTDTRWAADRWTDKDDDDDDDRDAGTAGTKRRRTW 729
QY      632  TTTCTATTTTAGTGAAGCGGAGAACGCGGAGAGCTGTCGCGTGAAGAGCTGATTC 691
DB      728  KSRWKEKRDTRWDADADDATARDRRRRGGDAGAGKKTGKRRERDRATWDRTDAWWA 669
QY      692  CGTTTTCGAGCTGCTGCTGGTGGCCGGAATGAACACCTACAAACCTGATTTCAATG 751
DB      668  DAAMWTTTDTDDWDKDRRRKRGARRRRRTTARAAWDWTWKAWDWAKWDKWTRADRW 609
QY      752  CGATGTA 758
DB      608  RWAADTW 602

RESULT 10
AZ934336/c
LOCUS       AZ934336             588 bp      DNA          linear      GSS 24-APR-2001
DEFINITION  Bc_Ba0002013r B. japonicum BAC library Bradyrhizobium japonicum

```

genomic, genomic survey sequence.
 A2934336
 VERSION A2934336.1 GI:13776396
 GSS.
 SOURCE Bradyrhizobium japonicum
 ORGANISM Bradyrhizobium japonicum
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Bradyrhizobium.
 1 (bases 1 to 588)
 Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
 Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
 genome
 Genome Res. 11 {B}, 1434-1440 (2001)
 JOURNAL 21376150
 MEDLINE 11483585
 PUBMED
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Class: BAC ends
 High quality sequence stop: 534.
 Location/Qualifiers
 source 1..588
 /organism="Bradyrhizobium japonicum"
 /mol_type="genomic DNA"
 /strain="USDA110"
 /db_xref="taxon:375"
 /lab_host="E. coli"
 /clone_lib="B. japonicum BAC library"
 /note="Vector: pIndigo536; Site_1: HindIII"

ORIGIN

Query Match 3.6%; Score 43.4; DB 28; Length 588;
 Best Local Similarity 45.8%; Pred. No. 0.26;
 Matches 149; Conservative 0; Mismatches 176; Indels 0; Gaps 0

QY 656 AACAGCGGAGAGCTCTCGGTGAAGAGTGTATTCGGTTTGACCGTCTGCTGCTGGTGG 715
 Db 328 AGAACCGCCACGGCAATGACCGAGAAGATTTGTTGCGCTACTGCTCATCATCAGGCTCAATA 269
 QY 716 CCGGAATGAACCACTACAAACCTGATTTCAATGCGATGTACAGCATATTAGAAACGC 775
 Db 268 CCGGACAGAGTCCACACCGAGCTGTATGAGCAGCGGCTGTGCGCTGACCGTCAATC 209
 QY 776 CAGCGGTTTACGAGGAACCTCGCAGCGCATCTCTGAACACTGATSCCTCAGGCACTGGAGGAAG 835
 Db 208 CCGATCAGAGCACCGGCTCTCGGCCACCCAGACATGATCAAGACGGCGTGGAGTAGA 149
 QY 836 CCTTGGCTTTCAGAGCGCGCGCCCGGTTTTTGGGCGCATTTGCCAAGCGGATACGGAGA 895
 Db 148 TGCTGGCGCTACGAGAGTTCCAACAGCTGGGCAACCGCATGACCACGAAACGGGTTCGAGC 89
 QY 896 TCGGGGGCGCACCTGATTAAAGAGGTGATATGTTTTGGCGTTTTGTGGCATCGGCAATC 955
 Db 88 TCGCGCGGCTATGCTCGATGCCGCGACGTGCGGTGACGCTGTGCATCGGGCGGCCAACCC 29
 QY 956 GTGATGAAGCAAGTTTGACAGACC 980
 Db 28 GCGACCCCGGCGAGTTTCGCGACCC 4

RESULT 11
 BZ563995/c
 LOCUS
 DEFINITION pac82-164_4486.y2 pac82-164 Pseudomonas aeruginosa genomic clone
 pac82-164_4486, genomic survey sequence.
 ACCESSION BZ563995
 VERSION BZ563995.1
 KEYWORDS GSS.
 GI:27188931

SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
AUTHORS 1 (bases 1 to 1136)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1. .1136
location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_4486"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

ORIGIN
Query Match 3.6%; Score 42.8; DB 28; Length 1136;
Best Local Similarity 48.7%; Pred. No. 0.53;
Matches 113; Conservative 0; Mismatches 119; Indels 0; Gaps 0

239 CCATCATTAACATGACCGCGGAGCATACAAAATCCGTTCACTCGTGAACAAGCCT 298
Db
359 CCATGCTWCAGCGCCAGCGCGGACCCGAGACACAATGCGCGCGGTTCGCGCGCGT 300

299 TTACTCCGCGGTGATGAAGCAATGGGAACCGAGAATTCAGAATACAGATGAACATGA 358
Db
299 TCTCCCGAGCCACACGAGCGACTCCAGACGATATAGACGCGGTACGCAAGANTGT 240

359 TTCAAAATTTCAGGGCGCGAGTTTGACCTTGTTTCAGATTTTTCATACCCGCTTC 418
Db
239 TGGAGCCATGTCGCGGCGGAGAACACACCGACGTAATGGCGCGCTTCGCGATCCACTGA 180

419 CGGTTATTGTGATCTGAGCTGCTGGAGTGCCCTTCAGCGCATGGAACA 470
Db
179 CCATCGCGGTGATCTACGACTGCTGGGCATTCGCGAGCGCGCGGAACA 128

RESULT 12
BX381961/c
LOCUS
DEFINITION BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cdna
clone CS0DI0721F05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961.1
KEYWORDS GI:30453007
SOURCE EST.
ORGANISM Homo sapiens (humar)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
REFERENCE Full-length cDNA libraries and normalization
AUTHORS Unpublished (2001)
TITLE Contact: Genoscope
JOURNAL Genoscope - Centre National de Sequencage
COMMENT BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :

```

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI072CC03NP1.
Location/Qualifiers
1. .1201
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="CS0DI072YF05"
   /tissue_type="PLACENTA COT 25-NORMALIZED"
   /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
   /note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
             source
Query Match          3.6%; Score 42.8; DB :3; Length 1201;
Best Local Similarity 5.7%; Pred. No. 0.54;
Matches 44; Conservative 253; Mismatches 469; Indels 3; Gaps 1
ORIGIN

```

Query Match 3.6%; Score 42.8; DB 13; Length 1201;
Best Local Similarity 5.7%; Pred. No. 0.54;
Matches 44; Conservative 253; Mismatches 459; Indels 3; Gaps 1

bits of the pcwvspork1.6 vector. Library was normalized.

QY	370	CAGGGCGCGATGAGTTTGACCTTTGATTCACGATTTTTCATACCGCTTCGGGTATTGTG	429
Db	1088	CSKGGCSKGVYVQGBMMKKMMKSVMMGKMMMKBMKMMKKKKKKKKKKKKKKBBKV	1028
QY	430	ATATCTGAGCTGCTGGGAGTGCCTTCAGCGCAGATGGAAAGTTTAAAGCATGCTCTCAT	489
Db	1028	KKKKTVKKKBMKTGKKKKKKGGKMMGMBMKMMGKGMVHGKGBMMBGKGMKMMMTGK	969
QY	490	CTTCTGCTCAGTACACCGAAGGATAAAGTGAAGAAGCTGAAAAGCTTTTTCGAAGAA	549
Db	968	MVCKBMMGGGKMMKMBGMVKKTTMMMBKMMKMMKMMKMMKMMKMMKMMKMMKMM	909
QY	550	CGAGATAAGTGTGAGGAAGACTGCGCGCGTTTTCGCGCGCATCATAGAAAGAAACGGA	609
Db	908	MKGGGMMKMMKMMVKKMMKKKKKKKKKKKKMMKMMKMMKMMKMMKMMKMMKMMKMM	849
QY	610	AAC---AAACCGGAACAGGATATATTCTATTATTAGTGAAGCGGAAGAAAACAGCGCAG	666
Db	848	MMVKGMMKMGKKKVMKGGKMMMMKMMKMMKMMKMMKMMKMMKMMKMMKMMKMMKMM	789
QY	667	AAGCTGTCGGTGAAGAGCTGATTCGGTTTTCACGCTGCTGCTGGTGGCGGAAATGAA	726
Db	788	KMKMMNN	729
QY	727	ACCACTACAAACCTGATTTTCAAATCGATGTACAGCATATTAGAAACGCCAGCGGTTAC	786
Db	728	MNN	669
QY	787	GAGGAACTCGGCAGCACTCTGAACTGATGCTCAGGCAGTGGAGGAAGCCTTCGCTTTC	846
Db	668	NN	609
QY	847	AGAGCGCGCGCCCGGTTTTCGAGCGCATTCGCAAGCGGATACGGAGATCGGGGGGCAC	906
Db	608	NN	549
QY	907	CTGATTAAGAAGGTGATATGGTTTTCGGCTTTGTGGCATCGCAAAATCGTGATGAAGCA	966
Db	548	KKKMMNNNNNTKMMGNNNNKKMMKKKKMMKMMKMMGMMNNNNNNNNNNNNNNNN	489
QY	967	AAGTTTGACACCGCACATGTTTGATATCGCGCGCATCCCAATCCCGCATATTCGTTT	1026
Db	488	NMMNNNNKONGNN	429
QY	1027	GGCCACGCATCCATTTTTCGCTTGGGCGCGCTTCGCGCTTGGAAGCAAAATACCGG	1086
Db	428	KKMMMMKMMKMMNN	369
QY	1087	TTAAGCTCTTGATTTCTGCTTTTCTCTCATATGAGAGTCGCGCAGTATCA	1135
Db	368	GTMNNNNNTTMMNTTNCNNNTTCANTTCTKTTTNNNTTNNNNNTTCA	320

RESULT 12	
BX381961/c	
LOCUS	1201 bp mRNA linear EST 08-MAY-2003
DEFINITION	Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CSODI072YF05 3-PRIME, mRNA sequence.
ACCESSION	BX381961
VERSION	BX381961
KEYWORDS	BX381961.1 GI:30453007
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Li, W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :

RESULT 13

CG092084
LOCUS
DEFINITION
CG092084 ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTA0545F03,
genomic survey sequence.
ACCESSION
CG092084
VERSION
CG092084.1 GI:33974378
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 718)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.

TITLE

Maize Genomics Consortium

JOURNAL

Unpublished (2003)

COMMENT

Other GSSs: PUTBM26TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

FEATURES

source

1. .718
/location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBTA0545F03"
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

ORIGIN

Query Match 3.5%; Score 42.2; DB 29; Length 718;
Best Local Similarity 49.3%; Pred. No. 0.65;
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 866 TGAGCGGCGATTCCCAAGCGGGATACGAGATCGGGGGACCTGATTAAAGAGGTGATA 925
Db 13 TGGCGCGGAGCGGATGAGGACACATCGTTGGCGGTGAGGCGATCGCGAAGGGGGA 72
QY 926 TGGTTTGGCGTTTGCGATCGGCAAAATCGTGATGAAGCAAAAGTTTGACAGACCGCACA 985
Db 73 AGGTGCTGTGTGTATCATCTCCGCCAATCGCGAGAAAGCGCTTCCCGGACGCGATC 132
QY 986 TGTTTGATATCGCGCGCATCCCATCGCATATTCGGTTTGGCCACGCGATCCATT 1045
Db 133 GCTTCGAGCTGTGTCGAAGGCGCGGACATGTCGGCTTCGGGCGCGGCGACGCTCT 192
QY 1046 GCCTTGGGCGCCCGCTTGGCCGCTCTTGAAGCAAAATATCGCGTT 1088
Db 193 GGTCTGGATCGCGCTCGCGGAATGCGTATCGCGTT 235

RESULT 14

BUS77443
LOCUS
DEFINITION
BUS77443 432 bp mRNA linear EST 17-SEP-2002
sar79c08.y1 Gm-cl074 Glycine max cdna clone SOYBEAN CLONE ID:
Gm-cl074-8560 5' similar to SF:C901_ARATH Q42569 CYTOCHROME P450
90A1.1; mRNA sequence.

ACCESSION

BUS77443

VERSION

BUS77443.1 GI:23052952

KEYWORDS

EST.

SOURCE

Glycine max (soybean)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE

1 (bases 1 to 432)

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Coryell, V.,
Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R. and Wilson, R.

TITLE

Public Soybean EST Project

JOURNAL

Unpublished (1999)

COMMENT

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information

call: (800)-533-4363 or contact: ccu@resgen.com web site:

www.resgen.com

Seq primer: -40RP from Gibco.

Location/Qualifiers

1. .432

/organism="Glycine max"

/mol_type="mRNA"

/db_xref="taxon:3847"

/clone="SOYBEAN CLONE ID: Gm-cl074-8560"

/tissue_type="seedlings induced for HR (hypersensitive

response)"

/dev_stage="9-11 day old"

/lab_host="DH10B"

/clone_lib="Gm-cl074"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:

XhoI; The cDNA library was constructed from mRNA isolated

from 9-11 day old seedlings that were induced for HR

(hypersensitive response) by vacuum infiltrating plant

tissue with Pseudomonas syringae pv. glycinea carrying the

avrB gene (Genetics 141:1597-1604). Plant tissue (expanded

unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,

and 53 hrs after inoculation and their mRNA pooled equally

for cDNA construction. The library was prepared using the

Stratagene pBluescript II SK(+) library construction kit.

Complementary DNA was synthesized from mRNA using a primer

consisting of a poly(dT) sequence with an XhoI restriction

site. EcoRI adaptors were ligated to the blunt-ended cDNA

fragments followed by XhoI digestion. The cDNA insert is

protected from XhoI digestion via methylation during first

strand synthesis. The cDNA fragments were directionally

cloned into the EcoRI-XhoI restriction site of the

pBluescript vector. The ligated cDNA fragments were

transformed into E. coli ElectroMax DH10B host cells. Plant

care, inoculations, and library construction were

performed by Steve Clough (Lila Vodkin lab, University of

Illinois)."

ORIGIN

Query Match 3.4%; Score 40.6; DB 13; Length 432;
Best Local Similarity 49.3%; Pred. No. 1.6;
Matches 106; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 562 GAGGAAGAACTGGCGCGTTTTCGCGCATCATAGAGAAAGCGAAACCGGAA 621
Db 135 GTGGCAGAGCGGACTAGCGTTGCTATGAGGAGAGAGAGATGCTGTAACAAA 194
QY 622 CAGGATATATTTCTATTTTGTGAGCGGAGAGAAACAGCGAGAGCTGTCGGTGA 681
Db 195 GAGAAATGAGTGACATGCTTGGGGCACTGTGGGCTCGGCGGACCACTTCCGACG 254
QY 682 GAGCTGATCTCGGTTTTCACGCTGCTGCTGGCGGGAATGAACCACTCAACCTG 741


```

Db      255 GAAATAGTGGATTTCCTGCTTGGCTGCGCGGTAGGAAACACCTCTACCATA 314
Qy      742 ATTTCAATGCGATGTCAGCATATTAGAACGCC 776
Db      315 ATGACTCTTGGATCAAGTCTCTCACTGAGACTCC 349

RESULT 15
BX406481
LOCUS   BX406481 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION
ACCESSION BX406481
VERSION   BX406481.1 GI:30635544
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0AM012AA08QF1.
LOCATION/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM012YA15"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

FEATURES
source

ORIGIN

Query Match 3.4%; Score 40.6; DB 13; Length 1201;
Best Local Similarity 23.8%; Pred. No. 2.5;
Matches 54; Conservative 74; Mismatches 99; Indels 0; Gaps 0;

Qy      422 TTATGTGATATCTGAGTCTGCTGGAGTGCCTTCAGCGGAGATGGAACAGTTTAAAGCAT 481
Db      573 KKKKTCNNKKKKKKNNCTTKATNKKNKKTKNCCCANKKKKAAKAAKAAKAKKDC 632
Qy      482 GGTCTGATCTTCTGCTAGTACACCGAGGATAAAGTGAAGCTGAAGAGCTTTT 541
Db      633 CKADKKKKKKKKKKMAAAMAAAMAAAMAAAMAAAMAAAMAAAMAAAMAAAKKH 692
Qy      542 TGAAGACGAGATAAGTGTGAGGAGAACTGGCGCGGTTTTTTTTCGCGCATATAGAAG 601
Db      693 AKAAGKAAKKKMAKAAADKKKKTKAAMAAAMCAAAACCKKMKKKKAAKKAATAAKA 752
Qy      602 AAAAGCGAAACAAACCGGACAGGATATTATTCTATTATTAGTGAA 648
Db      753 AAAAKKAAAKAAKTAAMAAAMAAAMAAAMAAAMAAAMAAAMAAAMAAAKKAA 799

```

Search completed: May 28, 2004, 22:08:31
 Job time : 2413 secs